

J. Math. Biol. (2015) 71:595–646  
DOI 10.1007/s00285-014-0832-5

# Mathematical Biology



## On the eigenvalue effective size of structured populations

Ola Hössjer

Received: 15 December 2013 / Revised: 16 August 2014 / Published online: 18 September 2014  
© The Author(s) 2014. This article is published with open access at [Springerlink.com](http://Springerlink.com)

**Abstract** A general theory is developed for the eigenvalue effective size ( $N_{eE}$ ) of structured populations in which a gene with two alleles segregates in discrete time. Generalizing results of Ewens (Theor Popul Biol 21:373–378, 1982), we characterize  $N_{eE}$  in terms of the largest non-unit eigenvalue of the transition matrix of a Markov chain of allele frequencies. We use Perron–Frobenius Theorem to prove that the same eigenvalue appears in a linear recursion of predicted gene diversities between all pairs of subpopulations. Coalescence theory is employed in order to characterize this recursion, so that explicit novel expressions for  $N_{eE}$  can be derived. We then study  $N_{eE}$  asymptotically, when either the inverse size and/or the overall migration rate between subpopulations tend to zero. It is demonstrated that several previously known results can be deduced as special cases. In particular when the coalescence effective size  $N_{eC}$  exists, it is an asymptotic version of  $N_{eE}$  in the limit of large populations.

**Keywords** Eigenvalue effective size · Coalescence theory · Predicted gene diversity · Migration · Perron–Frobenius · Perturbation theory of eigenvalues

**Mathematics Subject Classification** 92D25 · 60F99

### 1 Introduction

The effective size  $N_e$  was introduced by Wright (1931, 1938) as the size of an ideal homogeneous population with the same rate of loss of heterozygosity per generation as the studied population. It has become one of the most important parameters

---

O. Hössjer (✉)

Division of Mathematical Statistics, Department of Mathematics, Stockholm University,  
Stockholm, Sweden  
e-mail: [ola@math.su.se](mailto:ola@math.su.se)

in population genetics and conservation biology, as reviewed for instance by [Crow and Denniston \(1988\)](#), [Orrive \(1993\)](#), [Caballero \(1994\)](#), [Wang and Caballero \(1999\)](#), [Waples \(2002\)](#) and [Charlesworth \(2009\)](#).

Several closely related variants of  $N_e$  exist, and [Crow \(1954\)](#) first distinguished between the inbreeding effective size  $N_{eI}$ , the quantity originally defined by Wright, and the variance effective size  $N_{eV}$ . He also introduced a random extinction parameter that quantifies the long term rate at which genetic variants are lost. It is equivalent to the eigenvalue effective size  $N_{eE}$ , defined in terms of the largest non-unit eigenvalue of a Markov chain of allele frequencies ([Ewens 1982, 2004](#)). The nucleotide diversity or mutation effective size  $N_{e\pi}$  is essentially the expected coalescence time of a pair of haploid individuals ([Ewens 1989; Durrett 2008](#)), whereas the coalescent effective size  $N_{eC}$  is defined for populations such that the ancestral tree of any finite number of individuals converges to a Kingman coalescent in the limit of large populations ([Kingman 1982; Nordborg and Krone 2002; Sjödin et al. 2005; Wakeley and Sargsyan 2009; Hössjer 2011](#)).

In this paper we provide a general theory of  $N_{eE}$  for mutation free structured populations, in which a selectively neutral marker (referred to as a gene) with two variants or alleles segregates. The population consists of  $s$  homogeneous subpopulations (geographic sites, age classes, sexes or combinations thereof) and evolves in discrete time, with constant census sizes of all subpopulations.

[Ewens \(2004\)](#) reviewed results on  $N_{eE}$  for homogeneous populations, showing that it agrees with  $N_{eI}$  and  $N_{eV}$  for the Wright-Fisher model ([Wright 1931; Fisher 1958](#)), Kimura's multi-hypergeometric model ([Kimura 1957](#)), conditional branching process models ([Karlin and McGregor 1965](#)) or more generally models in which offspring numbers are exchangeable ([Cannings 1974](#)).

Results on  $N_{eE}$  for structured populations are less complete. [Ewens \(1982\)](#) showed that  $N_{eE}$  may differ from  $N_{eI}$  and  $N_{eV}$  for two-sex models. [Caballero and Hill \(1992\)](#) and [Nagylaki \(1995\)](#) considered a number of diploid models and derived formulas for an effective size based on the long term decay of heterozygosity. [Chesser et al. \(1993\)](#) and [Wang \(1997a, b\)](#) analyzed the island model with two sexes. They derived linear recursion formulas for the inbreeding coefficient and the coancestry of individuals from the same and different subpopulations, and computed an effective size from the largest eigenvalue of this recursion. [Felsenstein \(1971\)](#) computed the effective size for models with  $s$  age classes, and found the effective size from the largest eigenvalue of a linear recursion of  $s^2$  non-identity by descent probabilities of genes drawn with replacement from all pairs of age groups. [Maruyama \(1970a\)](#) derived a similar effective number for the circular stepping stone model under large population and small migration rate limits. [Tufto et al. \(1996\)](#) and [Tufto and Hindar \(2003\)](#) defined the eigenvalue effective size from a linear recursion of covariances between of all pairs of subpopulations.

All these notions of effective size are derived in terms the largest eigenvalue  $\lambda$  of linear recursions of covariances or probabilities of identity by descent or state. [Whitlock and Barton \(1997\)](#) showed that these linear recursions are closely related. They also argued briefly that the transition matrix of the Markov chain of allele frequencies has its largest non-unit eigenvalue equal to  $\lambda$ , and therefore all effective sizes of the previous paragraph agree with  $N_{eE}$ .

Motivated by the argument of [Whitlock and Barton \(1997\)](#), our main purpose in this paper is to provide a general framework for exact and asymptotic computation of  $N_{eE}$  for a large class of structured populations with stochastic backward migration and exchangeable reproduction within subpopulations. In Sect. 2 we introduce the population genetic model and Ewens' definition of  $N_{eE}$  in terms of the rate at which the Markov process of allele frequencies in all subpopulations reaches an absorbing state, quantified by the largest non-unit eigenvalue  $\lambda$  of its transition matrix. In Sect. 3 we focus on gene diversities, i.e. probabilities of genes not being identical by state. We introduce an  $s^2$ -dimensional deterministic process of predicted gene diversities and prove that whenever it is a linear recursion,  $\lambda$  also equals the largest eigenvalue of the matrix  $A$  of this recursion. In Sect. 4 we show how the elements of  $A$  are obtained from coalescence theory in new settings that generalize previous work, as illustrated with several examples in Sect. 5. Asymptotic approximations of  $\lambda$  and  $N_{eE}$  are obtained from perturbation theory of eigenvalues of matrices in Sect. 6, when either the population gets large and/or the migration rate gets small. This gives novel asymptotic expressions for  $N_{eE}$  that, for instance, in the limit of large populations agrees with the coalescence effective size  $N_{eC}$  when the latter exists. A discussion follows in Sect. 7 and proofs are collected in the "Appendix".

## 2 Model of reproduction, migration and allele frequency change

Consider a population of  $N$  individuals, divided into  $s$  subpopulations

$$\mathcal{I} = \{1, \dots, s\} \quad (1)$$

of constant sizes  $N_1 = Nu_1, \dots, N_s = Nu_s$ , with  $u_i \geq 0$  and  $\sum_{i \in \mathcal{I}} u_i = 1$ . Each individual carries two copies of a selectively neutral gene so that subpopulation  $i$  has a total of  $2N_i$  genes.

The population evolves in discrete time (not necessarily generations)  $t = 0, 1, \dots$ , with the genes of each subpopulation  $k \in \mathcal{I}$  at time  $t - 1$  numbered  $g = 1, \dots, 2N_k$ , and  $v_{tkig}$  referring to the number of offspring of gene  $g$  that migrate to subpopulation  $i$  at time  $t$ . The total gene flow from  $k$  to  $i$  between  $t - 1$  and  $t$  is summarized by the backward migration rate

$$\mathcal{B}_{tik} = \frac{1}{2N_i} \sum_{g=1}^{2N_k} v_{tkig}, \quad (2)$$

i.e. the fraction of genes a time  $t$  and subpopulation  $i$  that originate from  $k$  at time  $t - 1$ . The matrix  $\mathcal{B}_t = (\mathcal{B}_{tij})$  is referred to as the *observed* backward migration matrix at time  $t$ . Since its row sums are one, it is the transition matrix of a Markov chain with state space  $\mathcal{I}$ .

Let  $\mathbf{v}_{tkg} = (v_{tk1g}, \dots, v_{tksg})$  summarize the frequency distribution of the offspring of gene  $g$  of subpopulation  $k$  at time  $t - 1$  in all subpopulations. Assume that  $\{\mathbf{v}_{tkg}\}_{g=1}^{2N_k}$  are exchangeable random vectors, conditionally on  $\mathcal{B}_t$ , and that  $\{\mathcal{B}_t\}$  are independent and identically distributed random matrices with

$$E(\mathcal{B}_t) = \mathbf{B}, \quad (3)$$

where  $\mathbf{B} = (B_{ik})$ , the expected backward migration matrix, is the transition matrix of a Markov chain with state space  $\mathcal{I}$ . We assume that the  $\mathbf{B}$  is irreducible and aperiodic, with a unique equilibrium distribution  $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_s)$  satisfying  $\gamma_i \geq 0$ ,  $\sum_{i=1}^s \gamma_i = 1$  and

$$\boldsymbol{\gamma} = \boldsymbol{\gamma} \mathbf{B}. \quad (4)$$

It follows from (2) that the observed forward migration rate between  $k$  and  $i$ , i.e. the expected number of offspring of the genes in subpopulation  $k$  at time  $t - 1$  that at time  $t$  end up in subpopulation  $i$ , is

$$\mathcal{M}_{tki} := E(v_{tkig} | \mathcal{B}_t) = \frac{u_i \mathcal{B}_{tik}}{u_k}. \quad (5)$$

In order to keep subpopulation sizes constant over time, the total contribution in (5) from all parental populations  $k$  must be constant, i.e.

$$u_i = \sum_{k=1}^s u_k \mathcal{M}_{tki}. \quad (6)$$

Let  $\mathcal{M}_t = (\mathcal{M}_{tki})$  be the observed forward migration matrix of time  $t$ . It follows from (3) and (5) that the corresponding expected forward migration matrix  $\mathbf{M} = E(\mathcal{M}_t)$  has elements  $M_{ki} = E(v_{tkig})$  related to those of  $\mathbf{B}$  as

$$B_{ik} = \frac{u_k M_{ki}}{u_i} \quad (7)$$

for  $1 \leq k, i \leq s$ . Taking expectations on both sides of (6), we find that the vector  $\mathbf{u} = (u_1, \dots, u_s)$  of relative subpopulation proportions satisfies

$$\mathbf{u} = \mathbf{u} \mathbf{M}; \quad (8)$$

a left eigenvector of  $\mathbf{M}$  with eigenvalue 1. The two vectors  $\mathbf{u}$  and  $\boldsymbol{\gamma}$  are identical for conservative migration (Nagylaki 1980), but in general they differ.

Consider a biallelic genetic marker, and let  $\mathbf{X}_t = (X_{t1}, \dots, X_{ts})'$  be a column vector of (relative) frequencies of one of the two alleles in all subpopulations at time  $t$ , where prime denotes transposition. Since  $\{v_{tkg}\}_{g=1}^{2N_k}$  are exchangeable, we may number the genes of subpopulation  $k$  and time  $t - 1$  so that the first  $2N_k X_{t-1,k}$  have the specified allele. Then the allele frequency drift from one time point to the next can be summarized as

$$X_{ti} = \frac{1}{2N_i} \sum_{k=1}^s \sum_{g=1}^{2N_k X_{t-1,k}} v_{tkig}. \quad (9)$$

The following result is a simple consequence of (3), (5) and (9):

**Proposition 1** Assume that  $\mathcal{B}_t$  is independent of  $\{X_s\}_{s \leq t-1}$ . Then the sequence  $\{X_t\}$  of allele frequencies satisfies

$$E(X_t | X_{t-1} = \mathbf{x}) = \mathbf{B}\mathbf{x}, \quad (10)$$

where  $\mathbf{x}$  is a column vector of allele frequencies of length  $s$ .

We can rephrase Proposition 1 as  $\{X_t\}$  being a vector-valued autoregressive process of order 1 (Brockwell and Davis 1991). This process will be heteroscedastic, since the covariance matrix  $\text{Var}(X_t | X_{t-1} = \mathbf{x})$  varies with  $\mathbf{x}$ . The dynamics of  $\{X_t; t \geq 0\}$  is described more generally by means of a time homogeneous Markov chain with a state space

$$\mathcal{X} = \left\{0, \frac{1}{2N_1}, \frac{2}{2N_1}, \dots, 1\right\} \times \dots \times \left\{0, \frac{1}{2N_s}, \frac{2}{2N_s}, \dots, 1\right\}$$

of size  $|\mathcal{X}| = \prod_{i=1}^s (2N_i + 1)$ , and a transition kernel  $\mathbf{P} = (P(\mathbf{x}, \mathbf{y}))$ , with elements  $P(\mathbf{x}, \mathbf{y}) = P(X_t = \mathbf{y} | X_{t-1} = \mathbf{x})$  for all  $\mathbf{x}, \mathbf{y} \in \mathcal{X}$ . Since our model is free of mutations and no subpopulation is isolated, sooner or later one of the two alleles will be fixed in all subpopulations. This can be phrased as  $\{X_t\}$  having two absorbing states  $\mathbf{0} = (0, \dots, 0)$  and  $\mathbf{1} = (1, \dots, 1)$ , so that  $\mathbf{P}$  is reducible with two stationary distributions  $\pi_1(\mathbf{x}) = 1_{\{\mathbf{0}\}}(\mathbf{x})$  and  $\pi_2(\mathbf{x}) = 1_{\{\mathbf{1}\}}(\mathbf{x})$ , one for each of the absorbing states, with  $1_{\mathcal{Y}}(\mathbf{x})$  the indicator function of  $\mathcal{Y} \subset \mathcal{X}$ . Write  $\boldsymbol{\pi}_i = (\pi_i(\mathbf{x}); \mathbf{x} \in \mathcal{X})$  for the corresponding two row vectors of length  $|\mathcal{X}|$ . Since  $\boldsymbol{\pi}_i = \boldsymbol{\pi}_i \mathbf{P}$  for  $i = 1, 2$ , they are left eigenvectors of  $\mathbf{P}$  with eigenvalue 1. We divide  $\mathcal{X} = \cup_{i=1}^n \mathcal{X}_i$  into components  $\mathcal{X}_1 = \{\mathbf{0}\}$ ,  $\mathcal{X}_2 = \{\mathbf{1}\}$ ,  $\mathcal{X}_3, \dots, \mathcal{X}_n$  that induce a block form

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & \mathbf{0} & \dots & \mathbf{0} \\ 0 & 1 & \mathbf{0} & \dots & \mathbf{0} \\ P_{31} & P_{32} & P_{33} & \dots & \mathbf{0} \\ \vdots & & & \ddots & \vdots \\ P_{n1} & P_{n2} & P_{n3} & \dots & P_{nn} \end{pmatrix} \quad (11)$$

of the transition matrix, with zero blocks above the diagonal. For any function  $\phi : \mathcal{X} \rightarrow \mathbb{R}$ , let  $\phi = (\phi(\mathbf{x}); \mathbf{x} \in \mathcal{X})'$  be a column vector of function values. Then  $\mathbf{P}$  acts as an operator  $\phi \rightarrow \mathbf{P}\phi$  on  $\mathbb{R}^{\mathcal{X}}$ , as

$$(\mathbf{P}\phi)(\mathbf{x}) = \sum_{\mathbf{y} \in \mathcal{X}} P(\mathbf{x}, \mathbf{y})\phi(\mathbf{y}) = E_{\mathbf{x}}(\phi(X_1)),$$

where  $E_{\mathbf{x}}$  denotes expectation conditionally on  $X_0 = \mathbf{x}$ , cf. e.g. Norris (2008). In particular, the column vectors generated from

$$\begin{aligned} \phi_1(\mathbf{x}) &= 1 - \boldsymbol{\gamma}\mathbf{x}, \\ \phi_2(\mathbf{x}) &= \boldsymbol{\gamma}\mathbf{x} \end{aligned} \quad (12)$$

are both right eigenvectors of  $\mathbf{P}$  with eigenvalue 1, i.e.  $\phi_i = \mathbf{P}\phi_i$  for  $i = 1, 2$ . Indeed, this follows from (4) and (10), since

$$(\mathbf{P}\phi_2)(\mathbf{x}) = E_{\mathbf{x}}(\gamma \mathbf{X}_1) = \gamma E_{\mathbf{x}}(\mathbf{X}_1) = \gamma \mathbf{B}\mathbf{x} = \gamma \mathbf{x} = \phi_2(\mathbf{x}),$$

and similarly for  $\phi_1$ .

In order to find the rate of fixation of one of the two alleles, we need to look at  $\mathbf{P}^t$  for large  $t$ . We apply Markov chain theory and find this rate among all possibly complex-valued eigenvalue of  $\mathbf{P}$ , as the largest non-unit one. More specifically, in the “Appendix” we use Perron–Frobenius Theorem (see for instance Cox and Miller 1965) as a main ingredient for establishing the following:

**Theorem 1** *Suppose the square submatrices  $\mathbf{P}_{ii}$  in (11) along the diagonal are irreducible and aperiodic with at least one row sum less than one, for  $i = 3, \dots, n$ . Then the eigenvalues  $\lambda_i = \lambda_i(\mathbf{P})$  of  $\mathbf{P}$  (including multiplicity), can be ordered as*

$$1 = \lambda_1 = \lambda_2 > \lambda_3 \geq |\lambda_4| \geq \dots \geq |\lambda_{|\mathcal{X}|}| \geq 0, \quad (13)$$

with

$$\lambda_3 = \max_{3 \leq i \leq n} \lambda_{\max}(\mathbf{P}_{ii}). \quad (14)$$

Moreover, if the maximum in (14) is attained uniquely for  $i = k$ , then

$$\mathbf{P}^t = \phi_1 \pi_1 + \phi_2 \pi_2 + \lambda_3^t \begin{pmatrix} 0 & 0 & \mathbf{0} & \dots & \mathbf{0} \\ 0 & 0 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{R}_{31} & \mathbf{R}_{32} & \mathbf{R}_{33} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_{n1} & \mathbf{R}_{n2} & \mathbf{R}_{n3} & \dots & \mathbf{R}_{nn} \end{pmatrix} + o(\lambda_3^t) \quad (15)$$

as  $t \rightarrow \infty$ , where  $\mathbf{R}_{ii} = \mathbf{0}$  for  $i \neq k$ ,  $\mathbf{R}_{kk} = \phi_k \mathbf{q}_k$ ,  $\mathbf{q}_k = (q_k(\mathbf{x}); \mathbf{x} \in \mathcal{X}_k)$  is a row vector and  $\phi_k = (\phi_k(\mathbf{x}); \mathbf{x} \in \mathcal{X}_k)'$  a column vector, both with strictly positive elements,  $\mathbf{R}_{ij}$  has non-negative elements for  $j \geq 3$ , and the remainder term is a matrix with all its  $|\mathcal{X}|^2$  elements of smaller order than  $\lambda_3^t$ .

The requirement on  $\mathbf{P}$  in Theorem 1 is very weak, essentially that  $\mathcal{X}_3, \dots, \mathcal{X}_n$  contain transient states, so that no subpopulation is isolated and eventually one of the two alleles will be fixed in all subpopulations. When  $m = 3$  there is only one component of transient states, and migration is then possible within a finite number of time steps, back and forth between any pair of subpopulations. A recursive formula is provided in the “Appendix” for all  $\mathbf{R}_{ij}$ , and for  $\mathbf{R}_{kk}$  we can normalize the two vectors  $\phi_k$  and  $\mathbf{q}_k$  so that  $\sum_{\mathbf{x} \in \mathcal{X}_k} q_k(\mathbf{x}) = \sum_{\mathbf{x} \in \mathcal{X}_k} \phi_k(\mathbf{x}) q_k(\mathbf{x}) = 1$ . Then  $\mathbf{q}_k$  is the quasi equilibrium distribution

$$q_k(\mathbf{x}) = \lim_{t \rightarrow \infty} P_{\pi}(\mathbf{X}_t = \mathbf{x} | \mathbf{X}_{t'} \in \mathcal{X}_k, \quad t' = 0, 1, \dots, t-1) > 0 \quad (16)$$

of  $X_t$  conditionally on starting and remaining in  $\mathcal{X}_k$  (Darroch and Seneta 1965; Collet and Martinez 2013). The quasi equilibrium distributions for the other  $\mathcal{X}_3, \dots, \mathcal{X}_n$  are part of the remainder term of (15).

The following important corollary of Theorem 1 deals with the asymptotic decay rate of the expected value of  $\phi(X_t)$  for a large class of functions:

**Corollary 1** *Let  $\phi : \mathcal{X} \rightarrow \mathbb{R}$  be a function satisfying*

$$\begin{aligned}\phi(\mathbf{x}) &= 0, \mathbf{x} \in \mathcal{X}_1 \cup \mathcal{X}_2 \\ \phi(\mathbf{x}) &\geq 0, \mathbf{x} \in \mathcal{X} \setminus (\mathcal{X}_1 \cup \mathcal{X}_2 \cup \mathcal{X}_k), \\ \phi(\mathbf{x}) &> 0, \mathbf{x} \in \mathcal{X}_k,\end{aligned}\tag{17}$$

where  $\mathcal{X}_k$  is the component for which the maximum in (14) is attained. Then

$$\begin{aligned}\lim_{t \rightarrow \infty} \frac{E_\pi(\phi(X_t))}{\lambda_3^t} &= \sum_{\mathbf{x}, \mathbf{y} \in \mathcal{X}} \pi(\mathbf{x}) \mathbf{R}(\mathbf{x}, \mathbf{y}) \phi(\mathbf{y}) \\ &\geq \sum_{\mathbf{x} \in \mathcal{X}_k} \pi(\mathbf{x}) \phi_k(\mathbf{x}) \cdot \sum_{\mathbf{y} \in \mathcal{X}_k} q_k(\mathbf{y}) \phi(\mathbf{y}),\end{aligned}\tag{18}$$

where  $\mathbf{R} = (\mathbf{R}_{ij}) = (R(\mathbf{x}, \mathbf{y}))$  is the matrix in (15), and  $E_\pi$  denotes expected value conditional on  $P_\pi(\mathbf{X}_0 = \mathbf{x}) = \pi(\mathbf{x})$ . In particular, the right hand side of (18) is strictly positive if  $\pi(\mathcal{X}_k) > 0$ .

Corollary 1 shows that the largest non-unit eigenvalue  $\lambda_3 = \lambda_3(\mathbf{P})$  determines the rate of decrease of the expected value  $E_\pi(\phi(X_t))$  as  $t \rightarrow \infty$ . Putting  $\phi(\mathbf{x}) = 1_{\{\mathbf{x} \notin \mathcal{X}_1 \cup \mathcal{X}_2\}}$ , we notice that the probability of non-fixation decreases with  $t$  at this rate, so that  $\lambda_3$  is the rate of fixation and the eigenvalue of main interest. We will often simplify notation and write

$$\lambda = \lambda_3(\mathbf{P}).\tag{19}$$

The Wright-Fisher (WF) model is a homogeneous population ( $s = 1$ ) of  $N$  diploid individuals, with  $X_t | X_{t-1} \sim \text{Bin}(2N, X_{t-1})/(2N)$ . Feller (1951) found all the eigenvalues of the transition matrix for the WF model, and in particular

$$\lambda_3(\mathbf{P}_{\text{WF}}) = 1 - \frac{1}{2N}.\tag{20}$$

For an allele frequency process  $\{X_t\}$  with transition matrix  $\mathbf{P}$ , we define the eigenvalue effective size

$$N_{eE} = \frac{1}{2(1 - \lambda)} = \frac{1}{2(1 - \lambda_3(\mathbf{P}))}\tag{21}$$

as the size of a WF population for which the largest non-unit eigenvalue in (20) is the same as for the studied population.

### 3 Rate of decay of predicted gene diversities

Following [Nei \(1973, 1987\)](#), we define the gene diversity

$$\mathcal{H}_{tij} = X_{ti}(1 - X_{tj}) + X_{tj}(1 - X_{ti}) \quad (22)$$

between subpopulations  $i$  and  $j$  at time  $t$  as the probability that two randomly chosen genes from  $i$  and  $j$  (picked with replacement if  $i = j$ ) are different by state, i.e. have different types of alleles. Regarding  $t = 0$  as present and  $t > 0$  as future, let

$$H_{tij} = E_{\pi}(\mathcal{H}_{tij}) \quad (23)$$

be the predicted gene diversity between  $i$  and  $j$  at time  $t$ , given an initial distribution  $\mathbf{X}_0 \sim \pi$ . We collect the predicted gene diversities between all pairs of subpopulations into a column vector

$$\mathbf{H}_t = \mathbf{H}_t(\pi) = \text{vec} \left( (H_{tij})_{i,j=1}^s \right) \quad (24)$$

of length  $s^2$ , where  $\text{vec}$  is the vectorization operator that converts a matrix into a column vector by stacking its columns on top of each other. In order to compute linear combinations of the elements of  $\mathbf{H}_t$ , we define weights  $W_{ij}$  for all pairs of subpopulations, and prove the following:

**Proposition 2** Suppose  $\mathbf{W} = \text{vec} \left( (W_{ij})_{i,j=1}^s \right)'$  is a row vector of length  $s^2$  with non-negative weights  $W_{ij} \geq 0$  satisfying the symmetry condition  $W_{ij} = W_{ji}$  for all  $i, j$ , and let  $\phi_{\mathbf{W}}(\mathbf{x}) = 2 \sum_{i,j=1}^s W_{ij} x_i (1 - x_j)$  be a quadratic functional of  $\mathbf{x} = (x_1, \dots, x_s)'$ . Then

$$\mathbf{W} \mathbf{H}_t = \sum_{i,j=1}^s W_{ij} H_{tij} = E_{\pi}(\phi_{\mathbf{W}}(\mathbf{X}_t)).$$

To see the importance of Proposition 2, we notice that a sufficient condition for  $\phi_{\mathbf{W}}$  to satisfy (17) is that all  $W_{ij} > 0$ . It therefore follows from Corollary 1 that we can find  $\lambda = \lambda_3(\mathbf{P})$  and hence also  $N_{eE}$  from the rate of decrease to 0 of linear combinations of  $H_{tij}$ . It is therefore of interest to study the time dynamics of  $\mathbf{H}_t$ , and we will assume that a non-negative square matrix  $\mathbf{A}$  of order  $s^2$  exists, so that  $\mathbf{H}_t$  satisfies the linear recursion

$$\mathbf{H}_t = \mathbf{A} \mathbf{H}_{t-1} \quad (25)$$

for  $t = 1, 2, \dots$ . It will be convenient to introduce

$$\mathcal{I}_2 = \mathcal{I} \times \mathcal{I}, \quad (26)$$



the set of all pairs of subpopulations (cf. (1)), and write the elements of  $\mathbf{A}$  as

$$\mathbf{A} = (A_{ij,kl})_{ij \in \mathcal{I}_2, kl \in \mathcal{I}_2}, \quad (27)$$

where  $ij$  and  $kl$  is short hand notation for the row and column numbers obtained from the stacking procedure of the  $\text{vec}$  operation. We can always divide (26) into  $m$  irreducible components  $\mathcal{I}_2 = \mathcal{C}_1 \cup \dots \cup \mathcal{C}_m$ . After a possible reordering of the elements of  $\mathcal{I}_2$ , this gives a corresponding block decomposition

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \dots & \mathbf{0} \\ \vdots & & \ddots & \vdots \\ \mathbf{A}_{m1} & \mathbf{A}_{m2} & \dots & \mathbf{A}_{mm} \end{pmatrix} \quad (28)$$

of (27), with  $\mathbf{A}_{aa} = (A_{ij,kl})_{ij,kl \in \mathcal{C}_a}$  irreducible for  $a = 1, \dots, m$ . Typically, the pair of ancestors of two genes, picked from any pair  $\mathcal{I}_2$  of subpopulations, will ultimately belong to  $\mathcal{C}_1$ , provided the ancestry is traced sufficiently far back in time. The following fundamental result clarifies the importance of  $\mathbf{A}$ :

**Theorem 2** *Let  $\{X_t\}$  satisfy the conditions of Theorem 1, with  $\lambda = \lambda_3(\mathbf{P})$  the largest non-unit eigenvalue of its transition matrix  $\mathbf{P}$ , defined in (19). Assume that (25) holds, with  $\mathbf{A}$  having non-negative elements. Then*

$$\lambda = \lambda_{\max}(\mathbf{A}) = \max_{1 \leq a \leq m} \lambda_{\max}(\mathbf{A}_{aa}), \quad (29)$$

where  $\mathbf{A}_{aa}$  are the diagonal matrices of (28). If the maximum in (29) is attained for a unique  $1 \leq c \leq m$ , then

$$\mathbf{A}^t = \lambda^t \mathbf{r} \boldsymbol{\rho} + o(\lambda^t) \text{ as } t \rightarrow \infty, \quad (30)$$

with  $\boldsymbol{\rho} = \text{vec}((\rho_{ij})_{ij \in \mathcal{I}_2})'$  and  $\mathbf{r} = \text{vec}((r_{ij})_{ij \in \mathcal{I}_2})$  left and right eigenvectors

$$\begin{aligned} \boldsymbol{\rho} \mathbf{A} &= \lambda \boldsymbol{\rho}, \\ \mathbf{A} \mathbf{r} &= \lambda \mathbf{r} \end{aligned} \quad (31)$$

of  $\mathbf{A}$  with eigenvalue  $\lambda$ . Explicit expressions for  $\boldsymbol{\rho}$  and  $\mathbf{r}$  are provided in the ‘‘Appendix’’, and their components can be normalized so that  $\rho_{ij}, r_{ij} > 0$  for  $ij \in \mathcal{C}_c$ ,  $\sum_{ij \in \mathcal{C}_c} \rho_{ij} = \sum_{ij \in \mathcal{C}_c} \rho_{ij} r_{ij} = 1$ ,  $\rho_{ij} = 0$  for  $ij \in \mathcal{C}_{c+1} \cup \dots \cup \mathcal{C}_m$  and  $r_{ij} = 0$  for  $ij \in \mathcal{C}_1 \cup \dots \cup \mathcal{C}_{c-1}$ .

The following key result follows from (21) and Theorems 1–2:

**Corollary 2** *Suppose  $\{X_t\}$  satisfies the conditions of Theorem 1, with a linear recursion (25) for predicted gene diversities in terms of a non-negative quadratic matrix*

A. Then the eigenvalue effective size is

$$N_{eE} = \frac{1}{2(1 - \lambda_{\max}(\mathbf{A}))}. \quad (32)$$

#### 4 Coalescence probabilities

In this section we derive a linear recursion (25) for the predicted gene diversity vector  $\mathbf{H}_t$ , using probabilities that ancestral lineages of two genes coalesce. This enables us to compute  $N_{eE}$  from (32).

Many authors have derived linear recursions for identity by descent probabilities, gene diversities, covariances or coalescence probabilities of subdivided populations with spatial, age, sex or some other structure, with or without mutations. This includes results in Malécot (1951), Hill (1972), Li (1976), Sawyer (1976), Felsenstein (1982), Slatkin (1991), Nagylaki (2000), Ryman and Leimar (2008), Durrett (2008), Hössjer and Ryman (2014), Hössjer et al. (2014) and other papers mentioned in the introduction. They utilize coalescence probabilities more or less explicitly. We will generalize several of these results for constant subpopulation sizes, allowing backward migration rates to be stochastic and reproduction within subpopulations to have a general form.

Consider a structured coalescent (Notohara 1990; Herbots 1997; Wakeley 1998) for two genes, with coalescence events formulated hierarchically in two steps, first for subpopulations and then for genes. We draw two different genes from the population at time  $t$  and consider their joint ancestral subpopulation history  $(I_\tau, J_\tau)_{\tau=0}^T$  at times  $\{t - \tau\}_{\tau=0}^T$  until they find their most recent common ancestor at  $t - T$ , where  $T$  is the coalescence time. Let

$$\mathcal{Q}_{t,ij,kl} = P((I_1, J_1) = (k, l) | (I_0, J_0) = (i, j), T > 0, \mathcal{B}_t) \quad (33)$$

be the probability that the parents of two genes drawn from subpopulations  $i$  and  $j$  at time  $t$  have parents from subpopulations  $k$  and  $l$ , conditionally on  $\mathcal{B}_t$ . Since there are  $2Nu_i \mathcal{B}_{tik}$  genes of subpopulation  $i$  at time  $t$  that originate from subpopulation  $k$ , it follows that

$$\mathcal{Q}_{t,ij,kl} = \begin{cases} \mathcal{B}_{tik} \mathcal{B}_{tjl}, & i \neq j, \\ 2Nu_i \mathcal{B}_{tik} (2Nu_i \mathcal{B}_{tik} - 1) / (2Nu_i (2Nu_i - 1)), & i = j, k = l, \\ 2Nu_i \mathcal{B}_{tik} \cdot 2Nu_i \mathcal{B}_{til} / (2Nu_i (2Nu_i - 1)), & i = j, k \neq l. \end{cases} \quad (34)$$

We gather all these probabilities into an observed backward migration matrix  $\mathbf{Q}_t = (\mathcal{Q}_{t,ij,kl})_{ij \in \mathcal{I}_2, kl \in \mathcal{I}_2}$  of order  $s^2$  for pairs of subpopulations. By averaging with respect to  $\mathcal{B}_t$ , we then define the unconditional probabilities

$$Q_{ij,kl} = E(\mathcal{Q}_{t,ij,kl}) = P((I_1, J_1) = (k, l) | (I_0, J_0) = (i, j), T > 0) \quad (35)$$

that the parents of two genes from  $i$  and  $j$  belong to  $k$  and  $l$ , and collect them into a square matrix  $\mathbf{Q} = (Q_{ij,kl})_{ij \in \mathcal{I}_2, kl \in \mathcal{I}_2}$  of order  $s^2$ . The following result summarizes the role of  $\mathbf{Q}_t$  and  $\mathbf{Q}$  and is stated without proof:

**Proposition 3** Suppose segregation is independent between generations and condition on that coalescence events do not take place. Then the joint subpopulation ancestry  $\{(I_\tau, J_\tau)\}_{\tau=0}^t$  of a pair of different genes from time  $t$  back to time 0 is a Markov chain with state space  $\mathcal{I}_2$  that (a) conditionally on  $\{\mathcal{B}_{t-\tau}; \tau \geq 0\}$  has time varying transition matrices  $\{\mathcal{Q}_{t-\tau}\}_{\tau=0}^{t-1}$ , (b) unconditionally has a time invariant transition matrix  $\mathcal{Q}$ .

The next step is to incorporate coalescence events. To this end, let

$$P_{t,ijk} = P(T = 1 | I_0 = i, J_0 = j, I_1 = J_1 = k, \mathcal{B}_t, T > 0) \quad (36)$$

be the probability that two genes from subpopulations  $i$  and  $j$  at time  $t$  have the same parent, given that the parent belongs to  $k$ , and conditionally on  $\mathcal{B}_t$ . The corresponding unconditional coalescence probability  $p_{ijk}$  that two genes from  $i$  and  $j$  that both have their parents in  $k$ , have the same parent, is

$$p_{ijk} = P(T = 1 | I_0 = i, J_0 = j, I_1 = J_1 = k, T > 0). \quad (37)$$

It will be helpful to introduce the quantities

$$V_{kij} = \begin{cases} E(v_{tki1}(v_{tki1} - 1)), & \text{if } i = j, \\ E(v_{tki1}v_{tkj1}), & \text{if } i \neq j, \end{cases} \quad (38)$$

in order to get a more explicit expression for  $p_{ijk}$ , where  $v_{tkig}$  in (2) is the number of offspring of gene  $g$  in subpopulation  $k$  at time  $t - 1$  that end up in  $i$  at time  $t$ . The variables  $\{V_{kij}\}_{i,j=1}^s$  quantify the average variability of reproductive success among individuals in subpopulation  $k$ , and the coalescence probabilities are closely related to standardized versions of them:

**Theorem 3** Suppose the backward subpopulation ancestry of two different genes before coalescence is a Markov chain, with transition matrix  $\mathcal{Q} = (Q_{ij,kl})$  as in (35), and define coalescence probabilities  $p_{ijk}$  as in (37). Then

$$p_{ijk} = \left( \frac{1}{1 - \frac{1}{2Nu_i}} \right)^{\{i=j\}} \frac{V_{kij}u_k}{2Nu_iu_jQ_{ij,kk}}, \quad (39)$$

with  $V_{kij}$  defined in (38), and the components (22) of the gene diversity vectors  $\mathcal{H}_t = \text{vec}((\mathcal{H}_{tij})_{i,j=1}^s)$  form a multivariate autoregressive process

$$\mathcal{H}_{tij} = \sum_{k,l=1}^s A_{ij,kl} \mathcal{H}_{t-1,kl} + \epsilon_{tij} \quad (40)$$

of order one for  $t = 0, 1, \dots$ , with  $\epsilon_{tij}$  satisfying  $E(\epsilon_{tij} | \mathcal{X}_{t-1}) = 0$ , and

$$A_{ij,kl} = \left( 1 - \frac{1}{2Nu_i} \right)^{\{i=j\}} \left( \frac{1 - p_{ijk}}{1 - \frac{1}{2Nu_k}} \right)^{\{k=l\}} Q_{ij,kl}. \quad (41)$$

The expected gene diversity  $\mathbf{H}_t$  satisfies (25), with  $\mathbf{A} = (A_{ij,kl})$  as in (41).

It turns out that the right eigenvector of  $\mathbf{P}$  corresponding to its third largest eigenvalue  $\lambda_3$  can be found explicitly, whenever (40) holds:

**Corollary 3** *Suppose the conditions of Theorems 1 and 2 hold, with a gene diversity process  $\mathcal{H}_t$  satisfying (40). Then the transition matrix  $\mathbf{P}$  of the Markov chain  $\mathbf{X}_t$  has a right eigenvector  $\phi_3 = (\phi_3(\mathbf{x}); \mathbf{x} \in \mathcal{X})'$ , corresponding to the third largest eigenvalue  $\lambda = \lambda_3(\mathbf{P})$ , with*

$$\phi_3(\mathbf{x}) = \sum_{i,j=1}^s \rho_{ij} (x_i(1-x_j) + x_j(1-x_i)), \quad (42)$$

where  $\rho = (\rho_{ij})$  is the left eigenvector (14) corresponding to the largest eigenvalue  $\lambda$  of  $\mathbf{A} = (A_{ij,kl})$ . Moreover,

$$\phi_3(\mathbf{x}) = \begin{cases} 0, & \mathbf{x} \in \mathcal{X}_1 \cup \dots \cup \mathcal{X}_{k-1}, \\ > 0, & \mathbf{x} \in \mathcal{X}_k, \\ \geq 0, & \mathbf{x} \in \mathcal{X}_{k+1} \cup \dots \cup \mathcal{X}_n, \end{cases}, \quad (43)$$

where  $\mathcal{X}_k$  is the component of  $\mathcal{X}_1 \cup \dots \cup \mathcal{X}_n$  for which the maximum in (14) is uniquely attained. In particular, the restriction of  $\phi_3$  to  $\mathcal{X}_k$  agrees (up to a multiplicative constant) with the function  $\phi_k$  of Theorem 1.

## 5 Examples

The key formula (41) provides a general way to find  $\mathbf{A}$  and hence also  $N_{eE}$  through (32). We study its two main ingredients; expected backward migration rates  $Q_{ij,kl}$  for pairs of genes in Subsect. 5.1, and coalescence probabilities  $p_{ijk}$  in Subsect. 5.2. Then we apply these findings to a number of models in Subsect. 5.3 in order to show the generality of Theorem 3, explain how to apply it and as a preparation for the asymptotics of Sect. 6.

### 5.1 Backward migration

*Example 1 (Fixed backward migration).* When the observed backward migration rates are non-random, we must have

$$\mathcal{B}_t = \mathbf{B} \quad (44)$$

in order to satisfy (3). It follows from (34), (35) and (44) that

$$Q_{ij,kl} = Q_{t ij,kl} = \begin{cases} B_{ik}B_{jl}, & i \neq j, \\ B_{ik}(2Nu_i B_{ik} - 1)/(2Nu_i - 1), & i = j, k = l, \\ 2Nu_i B_{ik}B_{il}/(2Nu_i - 1), & i = j, k \neq l. \end{cases} \quad (45)$$

Following the nomenclature of Sved and Latter ([Sved and Latter \(1977\)](#)), we refer to (44) as fixed migration rates.  $\square$

*Example 2 (Dirichlet multinomial backward migration).* Denote the  $i$ :th rows of  $\mathbf{B}$  and  $\check{\mathbf{B}}_t$  by  $\mathbf{B}_i = (B_{i1}, \dots, B_{is})$  and  $\check{\mathbf{B}}_{ti} = (\check{B}_{ti1}, \dots, \check{B}_{tis})$  respectively. Specify parameters  $0 < \alpha_i < \infty$ , a random matrix  $\check{\mathbf{B}}_t$  with rows

$$\check{\mathbf{B}}_{ti} \sim \text{Dir}(\alpha_i \mathbf{B}_i) \quad (46)$$

that are independent for  $i = 1, \dots, s$ , and assume

$$\mathbf{B}_{ti} | \check{\mathbf{B}}_{ti} \sim \frac{\text{Mult}(2Nu_i, \check{\mathbf{B}}_{ti})}{2Nu_i} \quad (47)$$

are conditionally independent and multinomially distributed random vectors, given all  $\check{\mathbf{B}}_{ti}$ . Combining (46) and (47), the rows of  $\mathbf{B}_t$  are independent random vectors with Dirichlet multinomial distributions. The parameters  $\alpha_i$  quantify the amount of variability of the rows of  $\check{\mathbf{B}}_t$ . We will also extend (46) to  $\alpha_i \in \{0, \infty\}$ , which in conjunction with (47) gives two degenerate cases within the Dirichlet multinomial family: When  $\alpha_i \equiv \infty$ ,  $\mathbf{B}_t = \mathbf{B}_i$  is fixed, and

$$\mathbf{B}_{ti} \sim \frac{\text{Mult}(2Nu_i, \mathbf{B}_i)}{2Nu_i}, \quad (48)$$

whereas if  $\alpha_i \equiv 0$  we have

$$\mathbf{B}_{ti} = \check{\mathbf{B}}_{ti} \sim \text{Mult}(1, \mathbf{B}_i), \quad (49)$$

so that when time proceeds backwards, the ancestral history of *all* genes within a subpopulation will vary according to the same Markov chain with transition matrix  $\mathbf{B}$ . From a forward perspective, the latter system has  $s$  subpopulation, but the backwards scenario will be identical to that of a single population whose size varies between  $2Nu_1, \dots, 2Nu_s$ , according to a Markov chain with transition kernel  $\mathbf{B}$ , see for instance [Jagers and Sagitov \(2004\)](#), [Sampson \(2006\)](#), [Pollak \(2010\)](#), [Kaj and Krone \(2003\)](#) and [Sano et al. \(2004\)](#).

For any  $0 \leq \alpha_1, \dots, \alpha_s \leq \infty$ , we notice that

$$E(\mathbf{B}_t) = E\left(E(\mathbf{B}_t | \check{\mathbf{B}}_t)\right) = E(\check{\mathbf{B}}_t) = \mathbf{B},$$

in accordance with (3). Since two genes of subpopulations  $i$  and  $j$  are drawn independently with multinomial distributions from rows  $i$  and  $j$  of  $\check{\mathbf{B}}_t$ , it follows from (34) that  $E\left(\mathcal{Q}_{tij,kl} | \check{\mathbf{B}}_t\right) = \check{B}_{tik} \check{B}_{tjl}$ . Since the rows of  $\check{\mathbf{B}}_t$  have independent Dirichlet distributions (46);

$$\begin{aligned}
Q_{ij,kl} &= E \left( E(Q_{ij,kl} | \check{\mathcal{B}}_t) \right) \\
&= E \left( \check{B}_{tik} \check{B}_{tjl} \right) \\
&= E(\check{B}_{tik}) E(\check{B}_{tjl}) + \text{Cov}(\check{B}_{tik}, \check{B}_{tjl}) \\
&= B_{ik} B_{jl} + 1_{\{i=j\}} \frac{1_{\{k=l\}} B_{ik} - B_{ik} B_{jl}}{\alpha_i + 1},
\end{aligned} \tag{50}$$

which simplifies to

$$Q_{ij,kl} = B_{ik} B_{jl} \tag{51}$$

when  $\alpha_i \equiv \infty$  and the rows of  $\mathcal{B}_t$  have multinomial distributions.  $\square$

## 5.2 Coalescence probabilities

*Example 3 (Mixed multinomial reproduction).* Coalescence probabilities require that a reproduction scheme is specified. A fairly general scheme is defined by dividing the time interval between  $t - 1$  and  $t$  into three steps. In a first fertilization step, a gamete pool of infinite size is created for each parental subpopulation  $k$ , to which the  $2N_k$  genes contribute in fractions  $\omega_{tk1}, \dots, \omega_{tk,2N_k}$  that are exchangeable random variables summing to one. In a second migration step, the  $s$  gamete pools mix, so that the infinitely sized post-migration gamete pool  $i$  is a mixture of pre-migration pools  $1, \dots, s$  in proportions  $\check{B}_{ti1}, \dots, \check{B}_{tis}$ . In the final reproduction step, subpopulation  $k$  at time  $t$  is formed by drawing  $2N_k$  genes from post migration gamete pool  $k$ . Then (47) holds, and regardless of the distribution of  $\check{B}_{tik}$ ,

$$(v_{tki1}, \dots, v_{tki,2N_k}) | \mathcal{B}_t, \omega_{tk} \sim \text{Mult}(2N_i \mathcal{B}_{tik}, \omega_{tk}) \tag{52}$$

independently between all pairs of parental and offspring subpopulations  $k$  and  $i$ , where  $\omega_{tk} = (\omega_{tk1}, \dots, \omega_{tk,2N_k})$ .

When (52) holds, we evaluate the expected value in (38) by conditioning on  $\mathcal{B}_{tik}$  and  $\omega_{tk}$  and then insert into (39). By second moment properties of the multinomial distribution; this yields

$$\begin{aligned}
p_{ijk} &= 2N_k E(\omega_{tk1}^2) \\
&= (2N_k)^{-1} + 2N_k \text{Var}(\omega_{tk1}) \\
&=: (2N_{eI,k})^{-1}
\end{aligned} \tag{53}$$

and this can also (more easily) be obtained by a direct argument. The coalescence probability  $p_{ijk} = p_k$  in (53) only depends on the parental subpopulation  $k$ , since  $\omega_{tk}$  is the same, regardless of the offsprings' subpopulations  $i$  and  $j$ . In the second step of (53) we used  $E(\omega_{tk1}) = 1/(2N_k)$ , since the components of  $\omega_{tk}$  are exchangeable. We interpret  $N_{eI,k} \leq N_k$  as a local inbreeding effective size of subpopulation  $k$ , with equality if and only if  $\omega_{tk}$  is non-random.  $\square$

*Example 4 (Survival indicators).* Consider the genes  $g = 1, \dots, N_k$  of subpopulation  $k$  at time  $t - 1$ , and assume that transitions  $k \rightarrow i$  from time  $t - 1$  to a specific subpopulation  $i$  at time  $t$  represent their survival, with  $v_{tkig} \in \{0, 1\}$  an indicator for gene  $g$  to be still alive at time  $t$ . Then

$$p_{iik} = 0 \quad (54)$$

follows from (38) to (39). For all other subpopulations  $j \neq i$  we let  $v_{tkjg}$  refer to the number of offspring of  $g$ . Then  $V_{kij} = P(v_{tki1} = 1)E(v_{tkj1} | v_{tki1} = 1) =: M_{ki}M_{k,j|i}$ , where  $M_{k,j|i}$  is the expected number of offspring in  $j$  for a gene that survives, possibly different from the expected number of offspring  $M_{kj} = E(v_{tkj1})$  of any gene in  $k$ . Since (34)–(35) imply  $Q_{ij,kk} = B_{ik}B_{jk}$ , it follows from (7) and (39) that the coalescence probability

$$p_{ijk} = \frac{1}{2Nu_k} \cdot \frac{M_{k,j|i}}{M_{kj}} \quad (55)$$

is the inverse of the local census size of  $k$  times a correction factor that quantifies how correlated survival is with number of progeny in  $j$ .  $\square$

### 5.3 Example models

*Example 5 (Cannings models).* For a homogeneous population ( $s = 1$ ), we may drop subpopulation index and write  $v_{t11g} = v_{tg}$ , with  $v_{t1}, \dots, v_{t,2N}$  exchangeable random variables. Since subpopulation sizes and backward migration probabilities are trivial ( $u_1 = Q_{11,11} = 1$ ), we only need to specify coalescence probabilities  $p = p_{111}$ . It follows from (38)–(39) that

$$p = \frac{E(v_{t1}(v_{t1} - 1))}{2N - 1},$$

and  $A = A_{11,11} = 1 - p$  due to (41). Insertion into (32) yields

$$N_{eE} = \frac{1}{2(1 - A)} = \frac{N - \frac{1}{2}}{E(v_{t1}(v_{t1} - 1))}, \quad (56)$$

in agreement with Section 3.7 of Ewens (2004). Notice that (56) implies  $N_{eE} = N$  for a Wright-Fisher model, since  $v_{t1} \sim \text{Bin}(2N, 1/(2N))$ .  $\square$

*Example 6 (Sink and source populations).* This population has a source of size  $N_1$  and a sink of size  $N_2$ . It is assumed that on average  $N_{\text{emig}}$  individuals emigrate from 1 to 2 per generation. The migration scheme is degenerate in the sense that the expected forward and backward migration matrices are both reducible;

$$\mathbf{M} = \begin{pmatrix} 1 & N_{\text{emig}}/N_1 \\ 0 & (N_2 - N_{\text{emig}})/N_2 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 1 & 0 \\ N_{\text{emig}}/N_2 & (N_2 - N_{\text{emig}})/N_2 \end{pmatrix},$$

although the equilibrium distribution  $\boldsymbol{y} = (1, 0)$  of  $\boldsymbol{B}$  is still unique. We assume multinomial backward migration (51), and reproduction where parents are drawn uniformly and independently from the parental subpopulation, corresponding to  $p_{ijk} = 1/(2N_k)$  in (53). If subpopulations are ordered as  $\mathcal{I}_2 = \{11, 12, 21, 22\}$ , we find from (41) that

$$\boldsymbol{A} = \begin{pmatrix} 1 - \frac{1}{2N_1} & 0 & 0 & 0 \\ B_{21} & B_{22} & 0 & 0 \\ B_{21} & 0 & B_{22} & 0 \\ (1 - \frac{1}{2N_2})B_{21}^2 & (1 - \frac{1}{2N_2})B_{21}B_{22} & (1 - \frac{1}{2N_2})B_{21}B_{22} & (1 - \frac{1}{2N_2})B_{22}^2 \end{pmatrix}.$$

It is evident that  $\boldsymbol{A}$  has a block structure (28), with  $\mathcal{C}_1 = 11$ ,  $\mathcal{C}_2 = 12$ ,  $\mathcal{C}_3 = 21$  and  $\mathcal{C}_4 = 22$ . Combining the expression for  $\boldsymbol{A}$  with (29) and (32), we obtain

$$N_{eE} = \frac{1}{2(1 - \max(1 - 1/(2N_1), B_{22}))} = \max\left(N_1, \frac{N_2}{2N_{\text{emig}}}\right).$$

For large enough migration rates  $N_{\text{emig}} \geq N_2/(2N_1)$ , the source population will determine the effective size, whereas for small migration rates  $N_{\text{emig}} < N_2/(2N_1)$ , the two populations get increasingly isolated, and the effective size tends to infinity. It turns out that Theorem 1 applies with  $n = 5$  components, of which the three non-absorbing ones are

$$\begin{aligned} \mathcal{X}_3 &= \{\boldsymbol{x} = (x_1, x_2); x_1 = 0, x_2 \neq 0\}, \\ \mathcal{X}_4 &= \{\boldsymbol{x} = (x_1, x_2); x_1 = 1, x_2 \neq 1\}, \\ \mathcal{X}_5 &= \{\boldsymbol{x} = (x_1, x_2); 0 < x_1, x_2 < 1\}. \end{aligned}$$

For large migration rates, the eigenvalue  $\lambda$  of  $\boldsymbol{P}$  and  $\boldsymbol{A}$  is found within components  $\mathcal{X}_5$  and  $\mathcal{C}_1$ , and (42) simplifies to  $\phi_3(\boldsymbol{x}) = \rho_{11}x_1(1 - x_1)$ . For small migration rates, we find  $\lambda$  within  $\mathcal{X}_3$ ,  $\mathcal{X}_4$  and  $\mathcal{C}_2$ ,  $\mathcal{C}_3$  respectively.  $\square$

*Example 7 (Combined age and spatial structure).* Age structured models have been studied by Felsenstein (1971), Hill (1972), Kaj et al. (2001), Sagitov and Jagers (2005) and Hössjer (2011). Here we consider a population with  $s = 2z$  that has two demes with  $z$  age classes each. Subpopulations are numbered so that  $i$  ( $z + i$ ) corresponds to age class  $i = 1, \dots, z$  of deme 1 (deme 2). The expected forward and backward migration matrices

$$\boldsymbol{M} = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}, \quad \boldsymbol{B} = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}, \quad (57)$$



have a block structure, with  $\mathbf{M}_{cd}$  and  $\mathbf{B}_{cd}$  describing migration between demes  $c$  and  $d$ . All blocks of the forward matrix have the same form, e.g.

$$\mathbf{M}_{11} = \begin{pmatrix} M_{11} & M_{12} & 0 & \dots & 0 & 0 \\ M_{21} & 0 & M_{2,3} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ M_{z-1,1} & 0 & 0 & \dots & 0 & M_{z-1,z} \\ M_{z,1} & 0 & 0 & \dots & 0 & 0 \end{pmatrix}, \quad (58)$$

with a first column containing expected number of offspring in deme 1 of all age classes in deme 1, and a superdiagonal with probabilities of surviving to the next age class *and* not migrating to deme 2. The blocks of the backward matrix a similar structure, for instance

$$\mathbf{B}_{11} = \begin{pmatrix} B_{11} & \dots & B_{1,z-1} & B_{1z} \\ B_{21} & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & B_{z,z-1} & 0 \end{pmatrix}, \quad (59)$$

with probabilities in the first row that the parent of a newborn in deme 1 originates from the various age classes in deme 1, and probabilities along the subdiagonal that genes of the various adult age classes in deme 1 did not migrate during the last time step. If parental subpopulations are chosen independently for all genes from  $\mathbf{B}$ , it follows that  $Q_{ij,kl}$  is given by (51).

Coalescence probabilities

$$p_{ijk} = \begin{cases} 1/(2N_{eI,k}), & a(i) = a(j) = 1, \\ 0, & a(i) = a(j) = a(k) + 1, \\ 1/(2N_k), & a(i) = 1, a(j) = a(k) + 1 \text{ or} \\ & a(i) = a(k) + 1, a(j) = 1, \end{cases} \quad (60)$$

are obtained for all triples  $ijk$  of age classes for which  $Q_{ij,kk}$  is nonzero, with  $a(i) = i \bmod z$  the age class of subpopulation  $i$ , assuming in the first row that parents of newborns in a particular deme are chosen from a mixed multinomial distribution, with coalescence probabilities as in (53). For the second row we used that the coalescence probability (54) for survival is zero, and in third row we assumed that survival is independent of number of offspring, cf. (55).

We finally obtain  $N_{eE}$  by inserting (57) and (59) into (51), and then (51) and (60) into (41) and (32).  $\square$

**Example 8 (Extended Moran model).** Eldon and Wakeley (2006, 2009) extended the Moran model to allow for more skewed offspring distributions, for a homogeneous population and for the island model. Here generalize their model to populations with conservative migration. The reproduction cycle between time points  $t - 1$  and  $t$  is divided into two parts. In the first reproduction step one gene within each subpopulation

$k$  at time  $t - 1$  is chosen randomly to have  $Y_{t-1,k} = Y_k$  offspring, including itself, where  $2 \leq Y_k \leq 2N_k$ . Then  $Y_k - 1$  other, randomly chosen genes from the same subpopulation die. In the second migration step we assume that (44) holds, so that the forward and backward migration rates are non-random and fixed. The conservative migration assumption implies that exactly  $2N_k$  genes from subpopulation  $k$  migrate in (randomly chosen) groups of sizes  $2N_1 B_{1k}, \dots, 2N_s B_{sk}$  to subpopulations  $1, \dots, s$ . The coalescence probability for any triple  $i, j, k$  of subpopulations is

$$p_{ijk} = p_k = \sum_{y=2}^{2N_k} P(Y_k = y) \frac{\binom{y}{2} \binom{2N_k - y}{0}}{\binom{2N_k}{2}} = \frac{E(Y_k(Y_k - 1))}{2N_k(2N_k - 1)}, \quad (61)$$

as shown either by a direct argument, or from (38), (39) and (45). We finally obtain  $N_{eE}$  by inserting (45) and (61) into (41) and (32).  $\square$

**Example 9 (Dioecious population).** Consider a population with  $N_m$  males,  $N_f = N - N_m$  females and sex ratio  $\xi = N_m/N$ . Inheritance at an autosomal locus is modeled with  $s = 4$  subpopulations; gametes within males inherited from the father ( $i = 1$ ) and mother ( $i = 2$ ), and gametes within females inherited from the father ( $i = 3$ ) and mother ( $i = 4$ ), so that the relative subpopulation sizes are  $\mathbf{u} = (\xi/2, \xi/2, (1 - \xi)/2, (1 - \xi)/2)$ . According to Mendelian laws, each gamete is inherited, with equal probability 0.5, either from a grandpaternal or a grandmaternal gamete within the father or mother. This gives an observed backward migration matrix  $\mathbf{B}_t$  with a multinomial distribution (48). In view of (7), the expected backward/forward migration matrices are

$$\mathbf{B} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}, \quad \mathbf{M} = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1-\xi}{2\xi} & 0 \\ \frac{1}{2} & 0 & \frac{1-\xi}{2\xi} & 0 \\ 0 & \frac{\xi}{2(1-\xi)} & 0 & \frac{1}{2} \\ 0 & \frac{\xi}{2(1-\xi)} & 0 & \frac{1}{2} \end{pmatrix}, \quad (62)$$

and  $\mathbf{y} = (1/4, 1/4, 1/4, 1/4)$  follows from (4). Pairwise backward migration probabilities  $Q_{ij,kl}$  are given by (51), and in order to find the coalescence probabilities  $p_{ijk}$ , we follow the notation of Hill (1979) and let  $m$  and  $f$  represent male and female sexes, write  $\tau_{rs}^2$  for the variance of the number of children of sex  $s$  of an individual of sex  $r$ , and  $\tau_{rr,rs}$  for the covariance of the number of children of sex  $r$  and  $s$  of an individual of sex  $r$ . It is shown in the ‘‘Appendix’’ that the nonzero coalescence probabilities when  $k = 1$  are

$$p_{111} = \frac{\tau_{mm}^2}{2Nu_1 - 1}, \quad (63)$$

$$p_{331} = \frac{1 + \frac{\xi^2}{(1-\xi)^2} \left( \tau_{mf}^2 - \frac{1-\xi}{\xi} \right)}{2Nu_1 - \xi/(1-\xi)}, \quad (64)$$

$$p_{131} = \frac{1 + \frac{\xi}{1-\xi} \tau_{mm,mf}}{2Nu_1}, \quad (65)$$

and analogous formulas hold for  $k = 2, 3, 4$ . In particular, if a sperm or ova gamete chooses its parent randomly among all  $N_m$  or  $N_f$  parental genes, the number of male and female offspring of a male are two independent binomial random variables  $\text{Bin}(2Nu_1, (2Nu_1)^{-1})$  and  $\text{Bin}(2Nu_3, (2Nu_1)^{-1})$ , with variances  $\tau_{mm}^2$  and  $\tau_{mf}^2$  respectively, and covariance  $\tau_{mm,mf} = 0$ . It is easily seen that in this case, all three probabilities in (63)–(65) equal  $1/(2Nu_1)$ . We finally obtain  $N_{eE}$  by inserting (62) into (51), and then (51), (63)–(65) and the analogous coalescence probabilities for  $k = 2, 3, 4$  into (41) and (32).  $\square$

## 6 Asymptotics

According to (32), we find  $N_{eE}$  from the largest eigenvalue  $\lambda$  of  $A$ , for which we derived very explicit expressions in Sect. 4. Here we will use this approach and give asymptotic formulas for  $N_{eE}$  when

$$A = A(\varepsilon) = (A_{ij,kl}(\varepsilon))_{ij,kl \in \mathcal{I}_2} \quad (66)$$

depends on a positive parameter  $\varepsilon \rightarrow 0$ , with a Taylor expansion

$$A(\varepsilon) = A(0) + \dot{A}\varepsilon + o(\varepsilon) \text{ as } \varepsilon \rightarrow 0, \quad (67)$$

and  $\dot{A} = (\dot{A}_{ij,kl})_{ij \in \mathcal{I}_2, kl \in \mathcal{I}_2}$  a matrix of order  $s^2$ . For each fixed  $\varepsilon > 0$ ,  $A(\varepsilon)$  satisfies the conditions of Theorem 2, so that in particular its unique largest eigenvalue is  $\lambda = \lambda(\varepsilon)$ . The limiting matrix  $A(0)$  is assumed to have a largest, not necessarily unique, eigenvalue  $\lambda(0) = 1$ . As in Maruyama (1970a) and Nagylaki (1980, 1995), we use perturbation theory of matrices (Horn and Johnson 1985; Friswell 1996), in order to deduce

$$\lambda(\varepsilon) = 1 + \dot{\lambda}\varepsilon + o(\varepsilon) \text{ as } \varepsilon \rightarrow 0, \quad (68)$$

for some negative constant  $\dot{\lambda} < 0$ . Inserting (68) into (21), we find the asymptotic rate

$$N_{eE}(\varepsilon) = \frac{1}{-2\dot{\lambda}\varepsilon} + o(\varepsilon^{-1}) \text{ as } \varepsilon \rightarrow 0 \quad (69)$$

by which the eigenvalue effective size tends to infinity. The following result gives a very explicit formula for  $\dot{\lambda}$ , see for instance Aa et al. (2007) and references therein for a proof:

**Theorem 4** *Suppose the above conditions hold, with  $A(0)$  having a largest eigenvalue  $\lambda(0) = 1$  of multiplicity  $1 \leq v \leq s^2$ , with corresponding left and right eigenvectors  $\rho_1(0), \dots, \rho_v(0)$  and  $r_1(0), \dots, r_v(0)$ . If also the perturbed left eigenvectors  $\rho_\alpha(\varepsilon)$  and eigenvalues  $\lambda_\alpha(\varepsilon)$  are differentiable functions of  $\varepsilon$  at 0 for  $\alpha = 1, \dots, v$ , it holds that  $\lambda(\varepsilon) = \max_{\alpha=1, \dots, v} \lambda_\alpha(\varepsilon)$  for small enough  $\varepsilon > 0$ , and (68) is satisfied, with*

$$\dot{\lambda} = \lambda_{\max}(\dot{A}) \quad (70)$$

and  $\dot{\mathbf{A}} = (\dot{A}_{\alpha\beta})$  a  $v \times v$  matrix with entries

$$\dot{A}_{\alpha\beta} = \rho_{\alpha}(0) \dot{\mathbf{A}} \mathbf{r}_{\beta}(0). \quad (71)$$

In particular, suppose  $\mathbf{A}(0)$  is the transition matrix of a Markov chain with a unique equilibrium distribution  $\rho(0) = (\rho_{ij}(0))$ , the left eigenvector corresponding to  $\lambda(0) = 1$ , and a right eigenvector  $\mathbf{r}(0) = \mathbf{1}$ . Then  $v = 1$  and

$$\dot{\lambda} = \sum_{ij,kl} \rho_{ij}(0) \dot{A}_{ij,kl}. \quad (72)$$

In the following three subsections, the small perturbation parameter  $\varepsilon$  will either correspond to an inverse population size, a migration rate or both. We will use (41) and establish a Taylor expansion (67) for each case based on

$$A_{ij,kl}(\varepsilon) = \left(1 - \frac{1}{2N_i(\varepsilon)}\right)^{\{i=j\}} \left(\frac{1 - p_{ijk}(\varepsilon)}{1 - \frac{1}{2N_k(\varepsilon)}}\right)^{\{k=l\}} Q_{ij,kl}(\varepsilon), \quad (73)$$

when population sizes  $N_i(\varepsilon) = N(\varepsilon)u_i(\varepsilon)$ , backward migration rates  $Q_{ij,kl}(\varepsilon)$  and/or coalescence probabilities  $p_{ijk}(\varepsilon)$  depend on the perturbation parameter  $\varepsilon \rightarrow 0$ . The asymptotic expression for  $N_{eE}$  is then obtained from (69) and (70).

## 6.1 Large populations

We assume that the population size  $N$  tends to infinity, while the relative subpopulations sizes  $\mathbf{u}$ , forward and backward migration matrices  $\mathbf{M}$  and  $\mathbf{B}$  are kept fixed. Introduce

$$\varepsilon = \frac{1}{2N^{\beta}} \quad (74)$$

as a perturbation parameter, with  $0 < \beta \leq 1$  a fixed constant. In order to verify a Taylor expansion of  $\mathbf{A}(\varepsilon)$  in (73), we first consider the backward migration matrix  $\mathbf{Q}(\varepsilon) = (Q_{ij,kl}(\varepsilon)) = E(\mathbf{Q}_t(\varepsilon))$  for pairs of genes. It does not depend on  $\varepsilon$  for Dirichlet multinomial backward migration in (50), whereas for fixed backward migration (45) it does. In order to keep generality we assume that  $\mathbf{Q}(\varepsilon)$  may depend on  $\varepsilon$ , with a Taylor expansion

$$\mathbf{Q}(\varepsilon) = \mathbf{Q}(0) + \dot{\mathbf{Q}}\varepsilon + o(\varepsilon) \text{ as } \varepsilon \rightarrow 0 \quad (75)$$

for some matrix  $\dot{\mathbf{Q}} = (\dot{Q}_{ij,kl})$ . It will be seen though that  $\dot{\mathbf{Q}}$  does not influence the asymptotic behavior of  $N_{eE}$ . In contrast, the asymptotic behavior of the coalescence probabilities  $p_{ijk} = p_{ijk}(\varepsilon)$  is crucial and depends on how variable reproductive

success is between individuals that migrate from one subpopulation ( $k$ ) to other pairs of subpopulations ( $i, j$ ). The limits

$$\sigma_{ijk} = \lim_{N \rightarrow \infty} \frac{V_{kij} u_k^2}{(Nu_k)^{1-\beta} u_i u_j Q_{ij,kk}(0)} \quad (76)$$

are assumed to exist for all triples  $ijk$ , with  $V_{kij}$  defined in (38). It follows from (39), (74) and (76) that the coalescence probabilities admit Taylor expansions

$$p_{ijk}(\varepsilon) = \frac{\sigma_{ijk}}{u_k^\beta} \varepsilon + o(\varepsilon) := \dot{p}_{ijk} \varepsilon + o(\varepsilon) \text{ as } \varepsilon \rightarrow 0 \quad (77)$$

for all  $i, j, k$ . We refer to  $\dot{p}_{ijk}$  as the coalescence rate when two lines from  $i$  and  $j$  are merged into  $k$  and time is measured in units of  $\varepsilon^{-1} = 2N^\beta$  generations. The coalescence rate  $\sigma_{ijk}$  takes the size of the parental subpopulation  $k$  into account and measures time in units of  $2(Nu_k)^\beta$  instead. Taking the  $\varepsilon \rightarrow 0$  limit in (73), it follows from (75) and (77) that

$$A(0) = Q(0). \quad (78)$$

We assume that  $Q(0)$  is the transition kernel of a Markov chain that is not necessarily irreducible (it may contain some transient states), but has a unique equilibrium distribution  $\rho(0) = (\rho_{ij}(0))$ , which is also the left eigenvector of  $Q(0)$  corresponding to its unique largest eigenvalue  $\lambda(0) = 1$ . Hence formula (72) of Theorem 4 applies, and we obtain the following:

**Theorem 5** Define  $\varepsilon$  and  $A = A(\varepsilon)$  as in (73)–(74). Assume the population size  $N \rightarrow \infty$  so that (75) holds and the limits in (76) exist for some  $0 < \beta \leq 1$ . Then  $A(\varepsilon)$  satisfies Taylor expansion (67), with  $A(0)$  as in (78) and

$$\begin{aligned} \dot{A}_{ij,kl} = & -1_{\{k=l\}} \frac{\sigma_{ijk} Q_{ij,kk}(0)}{u_k^\beta} + \dot{Q}_{ij,kl} \\ & + 1_{\{\beta=1\}} \left( 1_{\{k=l\}} \frac{Q_{ij,kk}(0)}{u_k} - 1_{\{i=j\}} \frac{Q_{ii,kl}(0)}{u_i} \right). \end{aligned} \quad (79)$$

If the differentiability conditions on  $\lambda(\cdot)$  and  $\rho(\cdot)$  in Theorem 4 hold, then

$$N_{eE} = \frac{N^\beta}{C} + o(N^\beta) \text{ as } N \rightarrow \infty, \quad (80)$$

with

$$C = \sum_{ijk} \frac{\rho_{ij}(0) \sigma_{ijk} Q_{ij,kk}(0)}{u_k^\beta}. \quad (81)$$

Suppose  $I = I_\tau$  and  $J = J_\tau$  are the subpopulations of the ancestors, taken from the same generation  $t - \tau$  of two genes sampled at a fixed time  $t$ . Let  $K = I_{\tau+1}$  and  $L = J_{\tau+1}$  refer to the subpopulations of their two parents. Arguing as in Möhle (1998a), it will take many generations before coalescence in a large population, so that  $(I, J, K, L)$  will first attain its equilibrium distribution  $\rho_{ij}(0)Q_{ij,kl}(0)$ . Therefore, (81) can be interpreted as a coalescence rate

$$C = E(1_{\{K=L\}}\dot{p}_{IJK}) = E\left(1_{\{K=L\}}\frac{\sigma_{IJK}}{u_K^\beta}\right) \quad (82)$$

at equilibrium if time is counted in units of  $2N^\beta$ . In view of this, we may expect that for large populations,  $N_{eE}$  is asymptotically equivalent to the coalescence effective size  $N_{eC}$  whenever the latter exists. In a number of examples below, we will indeed verify that

$$N_{eC} = \frac{N^\beta}{C}, \quad (83)$$

with  $C$  the same constant as in (81). To this end, we first need the following:

**Corollary 4** *Suppose the conditions of Theorem 5 hold. Then asymptotically as  $N \rightarrow \infty$ ,  $N_{eE}$  is given by (80)–(81), with*

$$Q_{ij,kl}(0) = B_{ik}B_{jl}, \quad (84)$$

for fixed backward migration (44), and

$$Q_{ij,kl}(0) = B_{ik}B_{jl} + 1_{\{i=j\}}\frac{1_{\{k=l\}}B_{ik} - B_{ik}B_{jl}}{\alpha_i + 1}, \quad (85)$$

for Dirichlet multinomial backward migration (46)–(47).

Theorem 2 and (85) imply that  $A$  has  $m = 1$  irreducible component  $\mathcal{C}_1 = \mathcal{I}_2$  for Dirichlet multinomial backward migration when  $B$  is irreducible and at least one  $\alpha_i > 0$ , whereas  $A$  has  $m = 2$  components  $\mathcal{C}_1 = \{(i, i); i = 1, \dots, s\}$  and  $\mathcal{C}_2 = \{(i, i); i \neq j\}$  when  $\alpha_i \equiv 0$ . Then the joint ancestry of two genes are confined to lie in the same subpopulation after a few generations, once their ancestral subpopulation lineages merge for the first time, although they may not yet have coalesced at the gene level.

**Corollary 5** *Assume the conditions of Theorem 5 hold, with Dirichlet multinomial backward migration and  $\alpha_i \equiv 0$ . Then the equilibrium distribution  $\rho(0) = (\rho_{ij}(0))$  of  $Q(0)$  is supported on the diagonal of  $\mathcal{I}_2$ , with elements*

$$\rho_{ij}(0) = 1_{\{i=j\}}\gamma_i. \quad (86)$$

Moreover,  $N_{eE}$  is asymptotically given by (80), with

$$C = \sum_{ik} \frac{\gamma_i \sigma_{iik} B_{ik}}{u_k^\beta}, \quad (87)$$

$$\sigma_{iik} = \lim_{N \rightarrow \infty} \frac{\bar{V}_{kii} u_k^2}{(Nu_k)^{1-\beta} u_i^2}, \quad (88)$$

$\bar{V}_{kii} = E(v_{tki1}(v_{tki1} - 1) | K_{ti} = k)$  and  $K_{ti}$  the subpopulation to which all parents of the genes in  $i$  at time  $t$  belong.

Jagers and Sagitov (2004) and Pollak (2010) studied populations with rapidly varying sizes, which can be viewed as a special case of the Dirichlet multinomial backward distribution with  $\alpha_i \equiv 0$  (see Example 2). They showed that  $N_{eC}$  satisfies (83), with  $\beta = 1$  and  $C$  as in (87).

It is also possible to get explicit expressions for  $C$  when backward migration is fixed or multinomial:

**Corollary 6** Assume the conditions of Theorem 5 hold, with fixed or multinomial ( $\alpha_i \equiv \infty$ ) backward migration. Then  $Q_{ij,kl}(0) = B_{ik} B_{jl}$  follows from (84) and (85) with  $\alpha_i \equiv \infty$  respectively, and the equilibrium distribution  $\rho(0) = (\rho_{ij}(0))$  has elements

$$\rho_{ij}(0) = \gamma_i \gamma_j. \quad (89)$$

Moreover,  $N_{eE}$  is asymptotically given by (80)–(81), with

$$C = \sum_{ijk} \frac{\gamma_i \gamma_j \sigma_{ijk} B_{ik} B_{jk}}{u_k^\beta}, \quad (90)$$

$$\sigma_{ijk} = \lim_{N \rightarrow \infty} \frac{V_{kij}}{(Nu_k)^{1-\beta} M_{ki} M_{kj}}, \quad (91)$$

and  $V_{kij}$  as defined in (38).

Felsenstein (1971) seems to have been first to use (89) for weighting pairs of subpopulations. Hössjer (2011) studied models with fixed backward migration, and showed that  $N_{eC}$  satisfies (83) when  $\beta = 1$ , with  $C$  as in (90).

**Example 10 (Local subpopulation sizes).** When the coalescence probability  $p_{ijk}(\varepsilon) = p_k(\varepsilon)$  only depends on the parental subpopulation  $k$  for all  $\varepsilon > 0$ , the size standardized coalescence rate (76) satisfies

$$\sigma_{ijk} = \sigma_k, \quad (92)$$

as for the mixed multinomial reproduction scheme of Example 3. We deduce from (53) that  $p_k \geq 1/(2Nu_k)$ , and

$$\sigma_{ijk} = \sigma_k = 1_{\{\beta=1\}} + 4 \lim_{N \rightarrow \infty} \left( (Nu_k)^{1+\beta} \text{Var}(\omega_{tk1}) \right), \quad (93)$$

with  $C$  larger and  $N_{eE}$  smaller the more variable the gene weights  $\omega_{tkg}$  are. When (92) and Theorem 5 hold, it follows that  $N_{eE}$  is given by (80), with

$$\begin{aligned} C &= \sum_k \frac{\sigma_k}{u_k^\beta} \sum_{ij} \rho_{ij}(0) Q_{ijkk}(0) \\ &= \sum_{k=1}^s \frac{\sigma_k \rho_{kk}(0)}{u_k^\beta}, \end{aligned} \quad (94)$$

using  $\sum_{ij} \rho_{ij}(0) Q_{ij,kl}(0) = \rho_{kl}(0)$  in the last step, since  $\rho(0)$  is the left eigenvector of  $Q(0)$  with eigenvalue 1. In particular, if all genes within each subpopulation have parents from the same subpopulation, (86) implies

$$C = \sum_{k=1}^s \frac{\sigma_k \gamma_k}{u_k^\beta}, \quad (95)$$

whereas for fixed or multinomial backward migration, we deduce

$$C = \sum_{k=1}^s \frac{\sigma_k \gamma_k^2}{u_k^\beta} \quad (96)$$

from (89). In particular, if offspring pick their parents uniformly and independently within the parental subpopulation  $k$ , we have  $p_{ijk} = p_k = 1/(2N_k)$ , so that the local inbreeding effective size  $N_{eI,k}$  in (53) equals the local census size  $N_k$ . Asymptotically, this corresponds to having  $\beta = 1$  in (74) and  $\sigma_k = 1$  in (92). For fixed backward migration, we can therefore use (80) and (96), and deduce  $N_{eE} = N/C + o(N)$  as  $N \rightarrow \infty$ , with

$$C = \sum_{k=1}^s \frac{\gamma_k^2}{u_k}. \quad (97)$$

Notohara (1993) and Nordborg and Krone (2002) showed that the coalescence effective size satisfies  $N_{eC} = N/C$ , with  $C$  as in (97). Whenever  $N_{eI,k} = N_k$ ,

$$\begin{aligned} \sum_{k=1}^s \frac{\gamma_k}{u_k} &= C(\alpha_i \equiv 0) \geq \\ \sum_{k=1}^s \frac{\gamma_k^2}{u_k} &= C(\alpha_i \equiv \infty) \geq \\ 1 &= C(\alpha_i \equiv \infty; \gamma_k = u_k). \end{aligned} \quad (98)$$

The first inequality of (98) shows how much stochastically varying migration lowers  $N_{eE}$  at most. Then Cauchy-Schwarz inequality shows how much a variable long



term reproductivity  $\gamma_k/u_k$  between subpopulations lowers  $N_{eE}$ , with equality for conservative migration  $\gamma_k = u_k$  (Nagylaki 1980). The circular stepping stone model has conservative migration with uniform population sizes  $u_k = 1/s$ , and Maruyama (1970a) found that  $N_{eE}/N \rightarrow 1$  as  $N \rightarrow \infty$ , in agreement with the right hand side of (98).  $\square$

*Example 11 (Multiple mergers).* Other limiting ancestral processes with multiple mergers (Pitman 1999; Sagitov 1999) are possible. Let  $p_{ijh,k}^{(3)}$  be the probability that three genes from subpopulations  $i, j, h$  that all have their parents in subpopulation  $k$ , have the same parent. We will only consider models for which the pairwise and triple coalescence probabilities  $p_{ijk} = p_k$  and  $p_{ijh,k}^{(3)} = p_k^{(3)}$  only depend on the source population. Then we must have

$$\frac{p_k^{(3)}(\varepsilon)}{p_k(\varepsilon)} \rightarrow 0 \text{ as } \varepsilon \rightarrow 0 \quad (99)$$

in order for the limiting process to be Kingman's coalescent, see Theorem 3.2 of Möhle (2000). It is not possible to violate (99) in Example 3. Indeed,

$$\frac{p_k^{(3)}}{p_k} = \frac{2Nu_k E(\omega_{tk1}^3)}{2Nu_k E(\omega_{tk1}^2)} = \frac{E(\omega_{tk1}^3)}{E(\omega_{tk1}^2)} \rightarrow 0,$$

as  $N \rightarrow \infty$ , since  $\omega_{tk1}$  is bounded by 1 and tends to zero in probability. On the other hand, the Moran model of Example 8 allows for multiple mergers for an appropriate choice of the offspring size  $Y_{t-1,k} = Y_k$  of the reproducing gene of subpopulation  $k$ . As in Eldon and Wakeley (2006), we let

$$P(Y_k = y) = \begin{cases} 1 - (2N_k)^{-\beta}, & y = 2, \\ (2N_k)^{-\beta}, & y = 2N_k\psi, \\ 0, & \text{otherwise,} \end{cases}$$

for some  $0 < \psi \leq 1$  and  $\beta > 0$ . Then one shows

$$p_k^{(3)} = \frac{E(Y_k(Y_k - 1)(Y_k - 2))}{2N_k(2N_k - 1)(2N_k - 2)}$$

analogously as in (61). Since

$$\begin{aligned} E(Y_k(Y_k - 1)) &\sim \psi^2(2N_k)^{2-\beta} + 2, \\ E(Y_k(Y_k - 1)(Y_k - 2)) &\sim \psi^3(2N_k)^{3-\beta}, \end{aligned}$$

as  $N \rightarrow \infty$ , it follows from (61) and the last two displayed equations that we can violate (99) when  $0 < \beta \leq 1$ , with  $\sigma_{ijk} = \sigma_k = \psi^2 2^{1-\beta}$ . Since the extended Moran model has fixed backward migration and conservative migration  $\gamma_k = u_k$ , it follows that  $N_{eE}$  satisfies (96) with

$$C = 2^{1-\beta} \psi^2 \sum_{k=1}^s u_k^{2-\beta}.$$

Notice that this expression equals 1 when  $\beta = \psi = 1$ , since then the coalescence probability is asymptotically equivalent to  $1/(2N_k)$ .  $\square$

Other applications of Theorem 5 with  $\beta = 1$  includes a single deme with  $s$  age classes. Explicit formulas for the constant  $C$  in  $N_{eE} = N/C + o(N)$  can be found under general assumptions on how reproductivity varies randomly between and within age classes, thereby extending results of Felsenstein (1971), Sagitov and Jagers (2005) and Hössjer (2011).

For dioecious models, Hill (1979) found that  $N_{eE} = N/C + o(N)$ , with

$$C = \frac{1}{16\xi} \left( 2 + \tau_{mm}^2 + 2 \frac{\xi}{1-\xi} \tau_{mm,mf} + \left( \frac{\xi}{1-\xi} \right)^2 \tau_{mf}^2 \right) + \frac{1}{16(1-\xi)} \left( 2 + \tau_{ff}^2 + 2 \frac{1-\xi}{\xi} \tau_{ff,fm} + \left( \frac{1-\xi}{\xi} \right)^2 \tau_{fm}^2 \right), \quad (100)$$

which also follows from (51), (62), (63)–(65) and (90). Other effective size expressions of a diploid population can be found in Crow and Denniston (1988), Caballero (1995) and Nagylaki (1995). The latter two authors also treat inheritance at sex-linked loci. The expression for  $C$  is then somewhat different, since males only have one copy of an X-chromosome, and only  $s = 3$  subpopulations are needed. Overlapping generations within a dioecious population (Pollak 2011) requires  $s = 4z$  ( $s = 3z$ ) subpopulations for inheritance at an autosomal (X-linked) locus with  $z$  age classes. See also Möhle (1998b), for coalescence theory of two-sex models.

## 6.2 Small migration rates

Assume that the subpopulations in (1) divide into  $m \leq s$  demes

$$\mathcal{I} = \mathcal{I}(1) \cup \dots \cup \mathcal{I}(m), \quad (101)$$

with deme  $d$  containing the subpopulations of  $\mathcal{I}(d)$ . We will introduce a migration parameter  $\varepsilon \rightarrow 0$  that quantifies the amount of migration between the demes (not within them) while the total population size  $N$  is kept fixed. In order to obtain an expression for  $N_{eE}$  as  $\varepsilon \rightarrow 0$ , the crucial part is to find how all  $Q_{ij,kl}(\varepsilon)$  in (73) depend on  $\varepsilon$ . Although the relative subpopulation sizes  $u_i(\varepsilon)$  and coalescence probabilities  $p_{ijk}(\varepsilon)$  may vary with  $\varepsilon$  to some extent, this will have no asymptotic impact on  $N_{eE}$  as  $\varepsilon \rightarrow 0$ .

We will assume that the backward migration matrix

$$\mathbf{B}(\varepsilon) = \mathbf{B}(0) + \varepsilon \dot{\mathbf{B}} \quad (102)$$

depends on  $0 \leq \varepsilon \leq \varepsilon_{\max}$ , where  $\varepsilon_{\max}$  is chosen to guarantee that  $\mathbf{B}(\varepsilon)$  remains a non-negative matrix. The demes are isolated when  $\varepsilon = 0$ , so that  $\mathbf{B}(0)$  has a block diagonal structure  $\mathbf{B}(0) = \text{diag}(\mathbf{B}_{11}(0), \dots, \mathbf{B}_{mm}(0))$ , with  $\mathbf{B}_{dd}(0) = (B_{ik}(0))_{i,k \in \mathcal{I}(d)}$  describing backward migration within deme  $d$ . Since  $\mathbf{B}(\varepsilon)$  is a transition matrix of a Markov chain for all  $\varepsilon$ , the row sums of  $\dot{\mathbf{B}}$  must be zero, and this holds, for instance, if

$$\dot{B}_{ik} = 1_{\{k \notin \mathcal{I}(d)\}} \dot{B}_{ik} - 1_{\{k \in \mathcal{I}(d)\}} B_{ik}(0) \sum_{l \notin \mathcal{I}(d)} \dot{B}_{il} \quad (103)$$

for all  $i \in \mathcal{I}(d)$  and  $d = 1, \dots, m$ . If  $\mathbf{M}(\varepsilon)$  and  $\mathbf{u}(\varepsilon)$  are computed for each  $\varepsilon > 0$  from (7) and (8), it follows from (102) that

$$\mathbf{M}(\varepsilon) = \varepsilon \dot{\mathbf{M}} + o(\varepsilon), \quad (104)$$

if all  $u_i(\varepsilon)$  are differentiable at 0, with  $\dot{\mathbf{M}} = (\dot{M}_{ki})$  having elements

$$\dot{M}_{ki} = u_i(0) \dot{B}_{ik}/u_k(0), \quad \text{when } k \in \mathcal{I}(a) \neq \mathcal{I}(b) \ni i. \quad (105)$$

The migration parameter  $\varepsilon$  is such  $\mathbf{B}(\varepsilon) = \dot{\mathbf{B}}\varepsilon + o(\varepsilon)$  and  $\mathbf{M}(\varepsilon) = \dot{\mathbf{M}}\varepsilon + o(\varepsilon)$  as  $\varepsilon \rightarrow 0$  for some positive constants  $\dot{\mathbf{B}}$  and  $\dot{\mathbf{M}}$ , where

$$B(\varepsilon) = \sum_{d=1}^m \sum_{i \in \mathcal{I}(d)} \gamma_i(\varepsilon) \sum_{k: k \neq \mathcal{I}(d)} B_{ik}(\varepsilon), \quad (106)$$

is the backward migration rate between demes, i.e. the average number of parents of ancestors far back in time that originate from another deme than their children, and

$$M(\varepsilon) = \sum_{d=1}^m \sum_{k \in \mathcal{I}(d)} u_k(\varepsilon) \sum_{i: i \neq \mathcal{I}(d)} M_{ki}(\varepsilon), \quad (107)$$

is the forward migration rate, i.e. the fraction of all offspring today whose parents reside in another deme. Backward migration  $B(\varepsilon)$  is somewhat easier to analyze theoretically, but often  $\mathbf{M}(\varepsilon)$  is of more interest in applications.

In order to find explicit expressions for  $\dot{\mathbf{B}}$  and  $\dot{\mathbf{M}}$ , we introduce  $\boldsymbol{\gamma}_d = (\gamma_{di})_{i \in \mathcal{I}(d)}$  as the equilibrium distribution of  $\mathbf{B}_{dd}(0)$ , and the matrix  $\mathbf{G} = (G_{ab})_{a,b=1}^m$  with elements

$$G_{ab} = \sum_{i \in \mathcal{I}(a)} \gamma_{ai} \sum_{k \in \mathcal{I}(b)} \dot{B}_{ik}. \quad (108)$$

It is the infinitesimal generator of a continuous time Markov process with state space  $\{1, \dots, m\}$  and an equilibrium distribution  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)$  satisfying

$$\begin{aligned}\boldsymbol{\theta} \mathbf{G} &= \mathbf{0}, \\ \sum_{d=1}^m \theta_d &= 1.\end{aligned}\tag{109}$$

In the next lemma we assume  $\varepsilon$  is small, so that migration is faster within than between demes, and subpopulations within a deme form a macro state. The backward ancestry of a gene then attains its equilibrium distribution within a deme before any transitions between demes occur, and then the backward deme ancestry is a continuous time Markov process with generator  $\mathbf{G}$ :

**Lemma 1** *Suppose  $\varepsilon \rightarrow 0$ . Then the equilibrium distribution  $\boldsymbol{\gamma}(\varepsilon)$  corresponding to (102) satisfies*

$$\gamma_i(\varepsilon) = 1_{\{i \in \mathcal{I}(d)\}} \theta_d \gamma_{di} + o(1),\tag{110}$$

for  $i = 1, \dots, s$ , and the backward migration rate (106)

$$B(\varepsilon) = -\varepsilon \sum_{d=1}^m \sum_{i,k \in \mathcal{I}(d)} \gamma_i(0) \dot{B}_{ik} + o(\varepsilon),\tag{111}$$

where  $\gamma_i(0)$  is the limit of the right hand side of (110). If all  $u_i(\varepsilon)$  are differentiable at 0, the forward rate (107) has a similar expansion

$$M(\varepsilon) = \varepsilon \sum_{d=1}^m \sum_{k \in \mathcal{I}(d)} u_k(0) \sum_{i \notin \mathcal{I}(d)} \dot{M}_{ki} + o(\varepsilon).\tag{112}$$

In order to find the asymptotic behaviour of  $N_{eE}$  as  $\varepsilon \rightarrow 0$  by means of (69) and Theorem 4, we derive an expression for  $\mathbf{A}(\varepsilon)$  in (73), find  $\dot{\mathbf{A}}$ , show that  $\mathbf{A}(0)$  has a largest eigenvalue  $\lambda(0) = 1$ , find its multiplicity  $\nu$  and corresponding left and right eigenvectors. Because all demes are isolated when  $\varepsilon = 0$ , it is easy to see that the ancestors of  $i \in \mathcal{I}(a)$  and  $j \in \mathcal{I}(b)$  must belong to  $k \in \mathcal{I}(a)$  and  $l \in \mathcal{I}(b)$  respectively. Therefore  $\mathbf{Q}(0)$  has a block diagonal structure

$$\mathbf{Q}(0) = \text{diag}(\mathbf{Q}_{ab}(0); 1 \leq a, b \leq m)\tag{113}$$

with  $\mathbf{Q}_{ab}(0) = (\mathbf{Q}_{ij,kl})_{i,k \in \mathcal{I}(a), j,l \in \mathcal{I}(b)}$  a square matrix of order  $|\mathcal{I}(a)||\mathcal{I}(b)|$  containing all backward transitions when one gene and its parent are from deme  $a$  and the other gene and its parent are from deme  $b$ . It follows from (73) that  $\mathbf{A}(0)$  has a block diagonal structure

$$\mathbf{A}(0) = \text{diag}(\mathbf{A}_{ab}(0); 1 \leq a, b \leq m)\tag{114}$$

as well, with  $A_{ab}(\varepsilon) = (A_{ij,kl}(\varepsilon))_{i,k \in \mathcal{I}(a), j, l \in \mathcal{I}(b)}$  having elements

$$\begin{aligned} A_{ij,kl}(\varepsilon) &= Q_{ij,kl}(\varepsilon) \\ &= B_{ik}(\varepsilon)B_{jl}(\varepsilon), \quad i, k \in \mathcal{I}(a), j, l \in \mathcal{I}(b), a \neq b, \end{aligned} \quad (115)$$

for any  $\varepsilon$  because of (73), for subpopulations  $i$  and  $j$  that reside in different demes. In particular,  $A_{ab}(0) = Q_{ab}(0)$  has a unique largest eigenvalue 1 when  $a \neq b$ , and associated left and right eigenvectors  $\rho_{ab}(0) = \text{vec}((\rho_{ab,ij}(0))_{ij \in \mathcal{I}_2})'$  and  $r_{ab}(0) = \text{vec}((r_{ab,ij}(0))_{ij \in \mathcal{I}_2})$  with components

$$\begin{aligned} \rho_{ab,ij}(0) &= \gamma_{ai}\gamma_{bj}1_{\{i \in \mathcal{I}(a), j \in \mathcal{I}(b)\}}, \\ r_{ab,ij}(0) &= 1_{\{i \in \mathcal{I}(a), j \in \mathcal{I}(b)\}}. \end{aligned} \quad (116)$$

Since coalescence events are possible within each deme, even when  $\varepsilon = 0$ , it follows that  $A_{dd}(0)$  differs from  $Q_{dd}(0)$ , with a largest eigenvalue strictly smaller than one. Therefore, the largest eigenvalue 1 of  $A(0)$  has multiplicity

$$v = |\{\alpha = (a, b); 1 \leq a \neq b \leq m\}| = m(m-1).$$

In order to apply Theorem 4 we must also find the entries of the matrix  $\dot{\Lambda} = (\dot{\Lambda}_{\alpha,\beta})_{\alpha,\beta=1}^v$ , where  $\alpha = (a, b)$ ,  $\beta = (c, d)$ ,  $a \neq b$  and  $c \neq d$ . Suppose  $i \in \mathcal{I}(a)$ ,  $j \in \mathcal{I}(b)$ ,  $k \in \mathcal{I}(c)$  and  $l \in \mathcal{I}(d)$ , then

$$\dot{A}_{ij,kl} = \dot{Q}_{ij,kl} = \begin{cases} B_{ik}(0)\dot{B}_{jl} + \dot{B}_{ik}B_{jl}(0), & c = a \neq b = d, \\ B_{ik}(0)\dot{B}_{jl}, & c = a \neq b \neq d, \\ \dot{B}_{ik}B_{jl}(0), & c \neq a \neq b = d, \\ 0, & c \neq a \neq b \neq d \end{cases} \quad (117)$$

follows from (114) and differentiation of (115) with respect to  $\varepsilon$ . Invoking the definition of  $\dot{\Lambda}$  in (71), we find that

$$\begin{aligned} \dot{\Lambda}_{ab,cd} &= \rho_{ab}(0)\dot{A}r_{cd} \\ &= \sum_{\substack{i,j \in \mathcal{I}(a) \times \mathcal{I}(b) \\ k,l \in \mathcal{I}(c) \times \mathcal{I}(d)}} \rho_{ab,ij}(0)\dot{A}_{ij,kl} \\ &= \sum_{\substack{i,j \in \mathcal{I}(a) \times \mathcal{I}(b) \\ k,l \in \mathcal{I}(c) \times \mathcal{I}(d)}} \gamma_{ai}\gamma_{bj}\dot{A}_{ij,kl}. \end{aligned} \quad (118)$$

Then we insert (117) into (118), make use of (108) and obtain

$$\dot{\Lambda}_{ab,cd} = \begin{cases} G_{aa} + G_{bb}, & c = a \neq b = d, \\ G_{bd}, & c = a \neq b \neq d, \\ G_{ac}, & c \neq a \neq b = d, \\ 0, & c \neq a \neq b \neq d. \end{cases} \quad (119)$$

In particular, when each subpopulation is a deme,  $m = s$  and  $\mathcal{I}(d) = \{d\}$  for  $d = 1, \dots, s$ , so that (108) implies  $\mathbf{G} = \dot{\mathbf{B}}$ , and  $\dot{\mathbf{\Lambda}} = (\dot{\Lambda}_{ij,kl})_{1 \leq i \neq j \leq s, 1 \leq k \neq l \leq s}$  is of order  $v = s(s-1)$ , with elements

$$\dot{\Lambda}_{ij,kl} = \begin{cases} \dot{B}_{ii} + \dot{B}_{jj}, & k = i \neq j = l, \\ \dot{B}_{jl}, & k = i \neq j \neq l, \\ \dot{B}_{ik}, & k \neq i \neq j = l, \\ 0, & \text{otherwise.} \end{cases} \quad (120)$$

Combining (111), (112) and (119) with Theorem 4, we obtain the following:

**Theorem 6** Suppose subpopulations are divided into  $m$  demes, as in (101), whose isolation is quantified by the matrix  $\dot{\mathbf{B}} = (\dot{B}_{ik})$  in (102). Then

$$N_{eE} = \frac{\sum_{d=1}^m \sum_{i,k \in \mathcal{I}(d)} \gamma_i(0) \dot{B}_{ik}}{2\lambda_{\max}(\dot{\mathbf{\Lambda}})} \cdot \frac{1}{B} + o(B^{-1}) \text{ as } B \rightarrow 0, \quad (121)$$

with  $B = B(\varepsilon)$  the backward migration rate between demes in (106), and  $\dot{\mathbf{\Lambda}} = (\dot{\Lambda}_{ab,cd})$  the matrix in (119), which simplifies to (120) when  $m = s$ . If all  $u_k(\varepsilon)$  are differentiable at  $\varepsilon = 0$  as well, then

$$N_{eE} = -\frac{\sum_{d=1}^m \sum_{k \in \mathcal{I}(d)} u_k(0) \sum_{i \notin \mathcal{I}(d)} \dot{M}_{ki}}{2\lambda_{\max}(\dot{\mathbf{\Lambda}})} \cdot \frac{1}{M} + o(M^{-1}) \text{ as } M \rightarrow 0, \quad (122)$$

with  $M = M(\varepsilon)$  the forward migration rate in (107) and  $\dot{M}_{ki}$  defined in (105).

**Example 12 (Island model).** The island model (Wright 1943; Maruyama 1970b) is the most well known example of a population with spatial substructure, having  $m = s$  demes, and a forward migration matrix

$$\mathbf{M}(\varepsilon) = (1 - \varepsilon)\mathbf{I} + \frac{\varepsilon}{s-1}(\mathbf{1}\mathbf{1}' - \mathbf{I}), \quad (123)$$

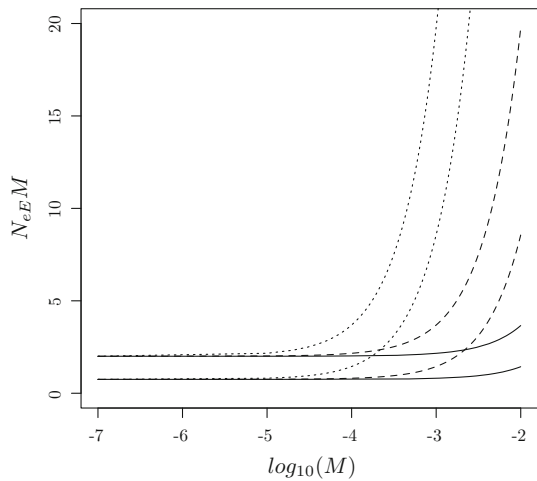
where  $\mathbf{1}$  is a column vector of  $s$  ones. Migration is symmetric, so that the migration rate  $M_{ki} = \varepsilon/(s-1)$  from each  $k$  to any other deme  $i \neq k$  is the same. It follows by symmetry from (4), (7) and (8) that  $\mathbf{B} = \mathbf{M}$ ,  $M = B = \varepsilon$  and  $u_k(\varepsilon) = \gamma_k(\varepsilon) = 1/s$ . This implies in particular that

$$\dot{B}_{ik} = \dot{M}_{ki} = \begin{cases} -1, & k = i, \\ 1/(s-1), & k \neq i. \end{cases} \quad (124)$$

Insertion of (124) into (120) yields

$$\dot{\Lambda}_{ij,kl} = \begin{cases} -2, & k = i \neq j = l, \\ 1/(s-1), & k = i \neq j \neq l, \\ 1/(s-1), & k \neq i \neq j = l, \\ 0, & \text{otherwise,} \end{cases}$$

**Fig. 1** Plots of  $N_{eE}M$  versus the migration rate  $M$  for the island model, with  $s = 4$  (lower curves) and  $s = 9$  (upper curves), when the local census size  $N/s$  equals 20 (solid), 200 (dashed) or 2,000 (dotted). The upper curves converge to  $(9 - 1)/4 = 2$  and the lower curves to  $(4 - 1)/4 = 3/4$  as  $M \rightarrow 0$ , in accordance with (125)



so that by symmetry, the largest eigenvalue of  $\dot{\mathbf{A}}$  corresponds to an eigenvector  $\mathbf{1}_v = (1, \dots, 1)'$  that is a column vector with  $v = s(s - 1)$  ones. Hence we find, from any of the row sums of  $\dot{\mathbf{A}}$ , that

$$\lambda_{\max}(\dot{\mathbf{A}}) = -2 + (s - 2) \cdot \frac{1}{s - 1} + (s - 2) \cdot \frac{1}{s - 1} = -\frac{2}{s - 1}.$$

We finally apply (122) and arrive at

$$\begin{aligned} N_{eE} &= \frac{\frac{1}{s} \sum_{i=1}^s (-1)}{2(-\frac{2}{s-1})} \cdot \frac{1}{M} + o(M^{-1}) \\ &= \frac{s-1}{4} \cdot \frac{1}{M} + o(M^{-1}) \end{aligned} \quad (125)$$

as  $M \rightarrow 0$ . The accuracy of this formula is illustrated in Fig. 1.  $\square$

**Example 13 (Circular stepping stone model).** The circular stepping stone model (Kimura 1953; Kimura and Weiss 1964; Maruyama 1970a) is a spatial model with  $m = s$  demes located along the perimeter of a circle, where migration from any deme is only possible to one of its two nearest neighbors. The elements of the expected forward migration matrix are

$$M_{ki}(\varepsilon) = \begin{cases} 1 - \varepsilon, & \text{if } k = i, \\ \varepsilon/2, & \text{if } \delta(k, i) = 1, \\ 0, & \text{otherwise,} \end{cases} \quad (126)$$

where  $\delta(i, j)$  is the shortest distance between demes  $i$  and  $j$  along the circle perimeter, when the distance between two neighboring demes is normalized to 1. It follows from

**Table 1** Comparison between the multiplicative constants  $N_{eE} = C/M + o(M^{-1})$  in (127), and the approximative multiplicative constant  $N_{eE} \approx C^{\text{appr}}/M + o(M^{-1})$  in (128), for a circular stepping stone model with  $s$  subpopulations with migration rate  $M \rightarrow 0$

$s$	$C$	$C^{\text{appr}}$
2	0.2500	0.2026
3	0.5000	0.4559
4	0.8536	0.8106
5	1.3090	1.2665
6	1.8660	1.8238
7	2.5245	2.4824
8	3.2843	3.2423
9	4.1454	4.1035
10	5.1079	5.0661

(4), (7) and (8) that  $\mathbf{M} = \mathbf{B}$ ,  $M = B = \varepsilon$  and  $u_k(\varepsilon) = \gamma_k(\varepsilon) = 1/s$ . Hence

$$\dot{M}_{ki} = \dot{B}_{ik} = \begin{cases} -1, & k = i, \\ 1/2, & \delta(k, i) = 1, \\ 0, & \text{otherwise,} \end{cases}$$

and from (120) we find that the matrix  $\dot{\mathbf{A}} = (\dot{A}_{ij,kl})$  has elements

$$\dot{A}_{ij,kl} = \begin{cases} -2, & k = i \neq j = l, \\ 1/2, & k = i \neq j, \delta(j, l) = 1, \\ 1/2, & \delta(k, i) = 1, i \neq j = l, \\ 0, & \text{otherwise.} \end{cases}$$

Since  $\sum_{k=1}^s u_k(0) \sum_{i;i \neq k} \dot{M}_{ki} = 1$ , we finally deduce from (122) that

$$N_{eE} = \frac{1}{-2\lambda_{\max}(\dot{\mathbf{A}})} \cdot \frac{1}{M} + o(M^{-1}) \text{ as } M \rightarrow 0. \quad (127)$$

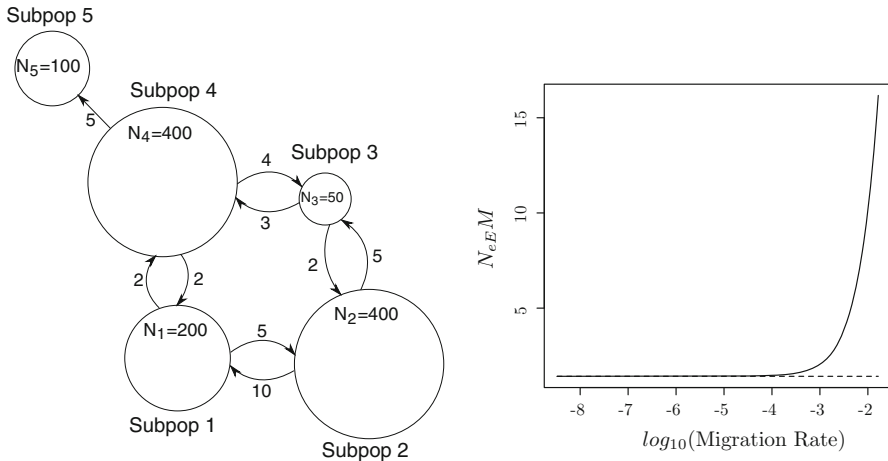
It seems difficult to obtain an explicit expression for the multiplicative constant in (127), although Maruyama (1970a) derived an approximation

$$N_{eE} \approx \frac{s^2}{2\pi^2} \cdot \frac{1}{M} + o(M^{-1}) \quad (128)$$

for even  $s$ . In Table 1 we compare (127) and (128) for different  $s$  and find a very good agreement.  $\square$

*Example 14 (System with five subpopulations).* A system with five subpopulations of varying size is shown in Fig. 2, with number of migrants in each generation depicted next to the arrows. The forward migration matrix is





**Fig. 2** Left System with five subpopulations of sizes  $N_i = Nu_i$  and, shown next to the arrows, the number of migrants  $N_k M_{ki}$  between different pairs  $k, i$  of subpopulations. Right Plots of  $N_{eE}M$  versus the migration rate  $M$  (solid curve) and the asymptotic limit  $C = 1.419$  (dashed line), calculated from the formula on the line below (131)

$$M = \begin{pmatrix} 0.94 & 0.025 & 0 & 0.01 & 0 \\ 0.025 & 0.9825 & 0.0125 & 0 & 0 \\ 0 & 0.04 & 0.82 & 0.06 & 0 \\ 0.005 & 0 & 0.01 & 0.9875 & 0.0125 \\ 0 & 0 & 0 & 0 & 0.95 \end{pmatrix},$$

and the relative subpopulation size vector  $\mathbf{u} = (2, 4, 0.5, 4, 1)/11.5$ . We let the forward migration rates depend on a perturbation parameter  $\varepsilon$  according to

$$M_{ki}(\varepsilon) = \begin{cases} \varepsilon M_{ki}, & i \neq k, \\ 1 + \varepsilon(M_{ii} - 1), & i = k, \end{cases} \quad (129)$$

so that  $\mathbf{u} = \mathbf{u}(\varepsilon)$  does not depend on  $\varepsilon$ , whereas the forward migration rate  $M(\varepsilon) = \varepsilon M$  is proportional to  $\varepsilon$ . It follows from (105) that

$$\dot{B}_{ik} = \begin{cases} \frac{u_k}{u_i} M_{ki}, & k \neq i, \\ (M_{ii} - 1), & k = i. \end{cases} \quad (130)$$

Combining (122) and (129), we find that

$$N_{eE} = \frac{C}{M} + o(M^{-1}) \text{ as } M \rightarrow 0, \quad (131)$$

with  $C = (1 - \sum_k u_k M_{kk})/(-2\lambda_{\max}(\dot{\mathbf{A}}))$  and  $\dot{\mathbf{A}}$  derived from (120) and (130). The numerically computed value  $C = 1.419$  is justified in Fig. 2.  $\square$

*Example 15 (Combined spatial and age structure).* Continuing Example 7, we assume that the forward migration matrix depends on  $\varepsilon$  as

$$\mathbf{M}(\varepsilon) = \mathbf{M}(0) + \varepsilon \dot{\mathbf{M}} = \begin{pmatrix} \mathbf{M}_{11}(0) - \varepsilon \dot{\mathbf{M}}_{12} & \varepsilon \dot{\mathbf{M}}_{12} \\ \varepsilon \dot{\mathbf{M}}_{21} & \mathbf{M}_{22}(0) - \varepsilon \dot{\mathbf{M}}_{21} \end{pmatrix},$$

so that the two demes are isolated when  $\varepsilon = 0$ . For brevity, write  $M_{ki} = M_{ki}(0)$ . The nonzero elements of the two off-diagonal blocks of  $\dot{\mathbf{M}}$  are

$$\begin{aligned} \dot{M}_{k,z+1} &= c_1 M_{k1} u_{z+1}, & k &= 1, \dots, s, \\ \dot{M}_{z+k,1} &= c_1 M_{z+k,z+1} u_1, & k &= 1, \dots, s, \\ \dot{M}_{k,z+k+1} &= c_{k+1} M_{k,k+1} u_{z+k+1}, & k &= 1, \dots, s-1, \\ \dot{M}_{z+k,k+1} &= c_{k+1} M_{z+k,z+k+1} u_{k+1}, & k &= 1, \dots, s-1, \end{aligned} \quad (132)$$

where  $c_1, \dots, c_z$  are non-negative constants, of which at least one is strictly positive. The migration rates in (132) are chosen so that  $u_k = u_k(\varepsilon)$  does not depend on  $\varepsilon$ . Intuitively, a fraction  $\varepsilon c_1 u_{z+1}$  of all offspring in deme 1 end up in deme 2, and a fraction  $\varepsilon c_{k+1} u_{z+k+1}$  of all genes of age class  $k$  of deme 1 that survive, migrate to deme 2, and similarly for the other two equations of (132). In the ‘‘Appendix’’ we verify that

$$N_{eE} = \frac{\sum_{i=1}^z c_i u_i u_{z+i}}{\sum_{i=1}^z c_i (\gamma_{1i} u_{z+i} + \gamma_{2i} u_i)} \cdot \frac{1}{M} + o(M^{-1}) \quad (133)$$

as  $M \rightarrow 0$ . When  $c_i$  increases with  $i$ , older individuals will migrate more, and this will increase  $N_{eE}$  if older individuals are less reproductive, and decrease  $N_{eE}$  if they reproduce more. Conservative migration is the intermediate case when all age groups are equally reproductive, with  $\gamma_{1i} = u_i/U(1)$ ,  $\gamma_{2i} = u_{z+i}/U(2)$  and  $U(d) = \sum_{i \in \mathcal{I}(d)} u_i$  the relative size of deme  $d$ . Insertion into (133) gives

$$N_{eE} = \frac{U(1)U(2)}{M} + o(M^{-1})$$

as  $M \rightarrow 0$  for conservative migration, independently of the age dependency of the migration pattern. In particular, if both demes are equally large, we get the same multiplicative constant  $C = U(1)U(2) = (1/2)^2 = 1/4$  as for an island model (125) with  $s = 2$ .  $\square$

### 6.3 Large populations and small migration rates

We let the inverse population size and the backward migration rate both tend to zero at the same speed, so that

$$\begin{aligned} \mathbf{B}(\varepsilon) &= \mathbf{B}(0) + \varepsilon \dot{\mathbf{B}}, \\ 4N\varepsilon &= c, \end{aligned} \quad (134)$$

with  $\varepsilon \rightarrow 0$  and  $c$  a constant. This can be viewed as an asymptotic scenario intermediate between (74) (with  $\beta = 1$ ) and (102).

The asymptotic expression for  $N_{eE}$  is derived similarly as in the previous subsection, so we only highlight the differences. Since the population size tends to infinity, the coalescence probabilities  $p_{ijk}$  will tend to zero, as described in (77), and this modifies (73) to

$$A_{ij,kl}(\varepsilon) = \left(1 - \frac{2\varepsilon}{cu_i}\right)^{\{i=j\}} \left(\frac{1 - \frac{2\sigma_{ijk}\varepsilon}{cu_k}}{1 - \frac{2\varepsilon}{cu_k}}\right)^{\{k=l\}} Q_{ij,kl}(\varepsilon) + o(\varepsilon), \quad (135)$$

where  $\sigma_{ijk}$  is the size standardized coalescence rate in (77). Consequently,

$$A(0) = Q(0) = \text{diag}(A_{ab}(0); 1 \leq a, b \leq m)$$

is a block diagonal matrix with blocks  $A_{ab}(0)$  given by (115) when  $\varepsilon = 0$  for all  $1 \leq a, b \leq m$ . These blocks have a unique largest eigenvalue 1, and

$$v = |\{\alpha = (a, b); 1 \leq a, b \leq m\}| = m^2.$$

When  $a \neq b$ , the left and right eigenvectors of  $A_{ab}(0)$  are as in (116). The same is true when  $a = b$  if we add the assumption of fixed or multinomial backward migration proportions. Differentiating (135) with respect to  $\varepsilon$  we find that

$$\dot{A}_{ij,kl} = \dot{Q}_{ij,kl} - \frac{2}{c} \left( 1_{\{k=l\}} \frac{\sigma_{ijk} - 1}{u_k} + \frac{1_{\{i=j\}}}{u_i} \right) Q_{ij,kl}(0), \quad (136)$$

with  $\dot{Q}_{ij,kl}$  as in (117), but without the restriction  $a \neq b$ . Therefore, inserting (116) and (136) into the definition of  $\dot{A} = (\dot{A}_{ab,cd})_{1 \leq a, b \leq m, 1 \leq c, d \leq m}$  in (71), we find, after some computations, that this matrix has elements

$$\dot{A}_{ab,cd} = \begin{cases} G_{aa} + G_{bb} - 1_{\{a=b\}} 2C_a/c, & c = a, d = b, \\ G_{bd}, & c = a, d \neq b, \\ G_{ac}, & c \neq a, d = b, \\ 0, & c \neq a, d \neq b, \end{cases} \quad (137)$$

with  $G_{ab}$  as in (108) and

$$C_a = \sum_{i,j,k \in \mathcal{I}(a)} \frac{\gamma_{ai} \gamma_{aj} \sigma_{ijk} B_{ik}(0) B_{jk}(0)}{u_k}$$

is a coalescence rate between the lines of deme  $a$  that can be interpreted as a local version of (90). In particular, when each subpopulation  $i$  is a deme,  $G_{ik} = \dot{B}_{ik}$ , and (137) reduces to

$$\dot{\Lambda}_{ij,kl} = \begin{cases} \dot{B}_{ii} + \dot{B}_{jj} - 1_{\{i=j\}} 2\sigma_{iii}/(cu_i), & k = i, l = j, \\ \dot{B}_{jl}, & k = i, l \neq j, \\ \dot{B}_{ik}, & k \neq i, l = j, \\ 0, & k \neq i, l \neq j. \end{cases} \quad (138)$$

Equipped with (137) and (138), we apply (69) and Theorem 4, and deduce:

**Proposition 4** *Suppose the migration rate between demes and the inverse population size tend to zero simultaneously as in (134) when  $\varepsilon \rightarrow 0$ , with a backward migration that is either fixed (44) or multinomial (48). The eigenvalue effective size then has an asymptotic expansion*

$$N_{eE} = -\frac{1}{2\lambda_{\max}(\dot{\Lambda})\varepsilon} + o(\varepsilon^{-1}), \quad (139)$$

with the elements of  $\dot{\Lambda}$  as in (137), or (138) when each deme contains one single subpopulation.

*Example 16 (Island model.)* We will assume that

$$\sigma_{iii} = \sigma \text{ for } i = 1, \dots, s, \quad (140)$$

where  $\sigma = (N/s)/N_{eI}$  can be interpreted as a ratio between the (constant) local census and effective size of each deme. This local inbreeding effective size  $N_{eI} = N_{eI,i}$  is similar to (53), although we only consider triplets  $i = j = k$  of demes here. It follows from (124), (138) and (140) that

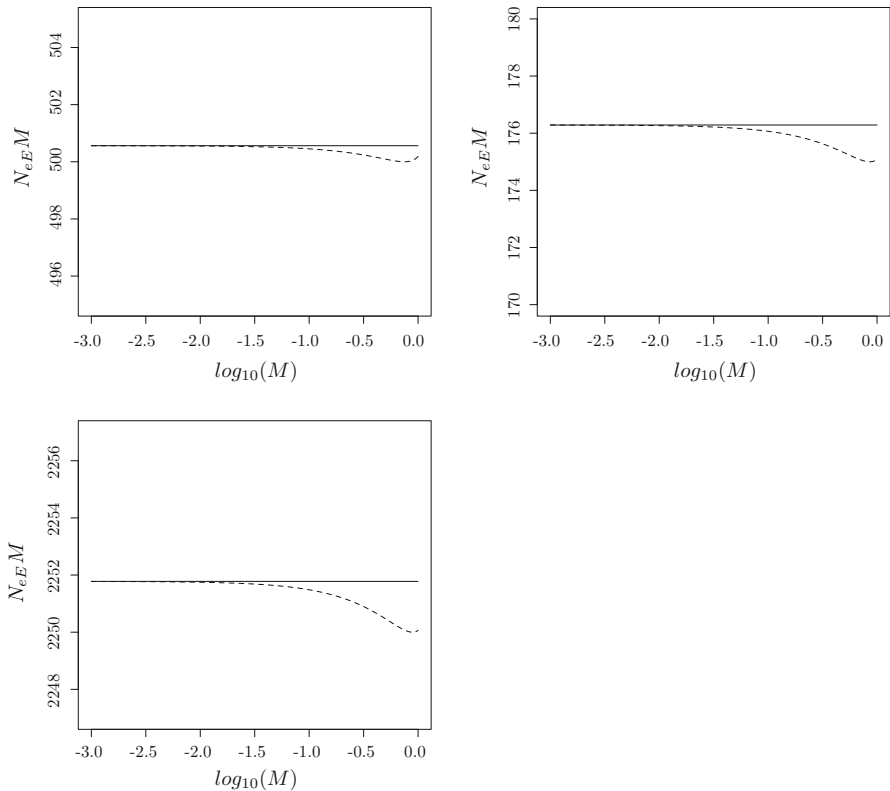
$$\dot{\Lambda}_{ij,kl} = \begin{cases} -2 - 1_{\{i=j\}} 2s\sigma/c, & k = i, l = j, \\ \frac{1}{s-1}, & k = i, l \neq j, \\ \frac{1}{s-1}, & k \neq i, l = j, \\ 0, & k \neq i, l \neq j. \end{cases} \quad (141)$$

Let  $\dot{\lambda} = \lambda_{\max}(\dot{\Lambda})$  be the largest eigenvalue of  $\dot{\Lambda}$ , with  $\mathbf{x} = \text{vec}\left((x_{ij})_{i,j=1}^s\right)$  the corresponding right eigenvector satisfying  $\dot{\Lambda}\mathbf{x} = \dot{\lambda}\mathbf{x}$ . By symmetry we must have  $x_{ij} = y$  when  $i = j$  and  $x_{ij} = z$  when  $i \neq j$ , and a  $2 \times 2$  system

$$\begin{cases} \dot{\lambda}y = (\dot{\Lambda}\mathbf{x})_{ii} = 2(z - y) - \frac{2s\sigma}{c}y \\ \dot{\lambda}z \stackrel{i \neq j}{=} (\dot{\Lambda}\mathbf{x})_{ij} = \frac{2}{s-1}(y - z) \end{cases} \quad (142)$$

of equations for  $y$  and  $z$ . It will be convenient to introduce the parameter  $\kappa = c/(s\sigma) = 4NM/(s\sigma) =: 4N_{eI}M$ . Then we apply (139) and find that

$$N_{eE} = \frac{1}{-2\dot{\lambda}(\kappa)M} + o(M^{-1}) \quad (143)$$



**Fig. 3** Plots of  $N_{eE}M$  versus  $M$  for the island model, when  $N_{eI} = N/s$  and  $\kappa = 4(N/s)M$  is kept constant, with  $s = 4$  and  $\kappa = 500$  (upper left),  $s = 7$  and  $\kappa = 100$  (upper right) and  $s = 9$  and  $\kappa = 1,000$  (lower left). The horizontal lines are the limits  $1/(-2\dot{\lambda}(\kappa))$  of  $N_{eE}M$  in (143) as  $M \rightarrow 0$

as  $M \rightarrow 0$ , where

$$\dot{\lambda}(\kappa) = -\left(\frac{s}{s-1} + \frac{1}{\kappa}\right) \left(1 - \sqrt{1 - \frac{4\kappa}{(s-1)\left(\frac{s\kappa}{s-1} + 1\right)^2}}\right)$$

is the largest root of the (quadratic) characteristic equation in  $\dot{\lambda}$  obtained from (142). Figure 3 verifies numerically fast convergence in (143), for three combinations of  $s$  and  $\kappa$  (notice the narrow scales of the y-axes). We have that  $\lim_{\kappa \rightarrow 0}(-2\dot{\lambda}(\kappa)) = 4/(s-1)$ , in agreement with (125). On the other hand,

$$N_{eE} = \frac{N}{C(\kappa)} + o(N^{-1}),$$

where

$$C(\kappa) = \sigma \cdot \left( \frac{s^2 \kappa}{2(s-1)} + \frac{s}{2} \right) \left( 1 - \sqrt{1 - \frac{4\kappa}{(s-1) \left( \frac{s\kappa}{s-1} + 1 \right)^2}} \right),$$

and when the migration rate dominates the inverse population size, we get  $\lim_{\kappa \rightarrow \infty} C(\kappa) = \sigma$  for  $\sigma_k = \sigma$  and  $\gamma_k = u_k = 1/s$ , in agreement with (96).  $\square$

## 7 Discussion

In this paper we developed a general theory which enables computation of the eigenvalue effective size  $N_{eE}$  for a large class of structured populations with stochastic backward migration and exchangeable reproduction within subpopulations, exactly or asymptotically when either the inverse population size and/or migration rates between subpopulations tend to zero.

Our work can be extended in several ways. First, subpopulation sizes could be time varying. Existence of  $\lambda$  then requires extra conditions, e.g. sizes that either vary as a Markov chain or cyclically. Several authors have studied this problem for homogeneous or age structured models, see Karlin (1968), Jagers and Sagitov (2004), Pollak (1980, 2002) and Wang and Pollak (2000a, b). For cyclically varying populations with period  $\tau$ , the matrix  $A = A_t$  of the predicted gene diversity recursion will depend cyclically on time. Whitlock and Barton (1997) argued that this deterministic process tends to zero at a rate

$$\lambda = \lambda_{\max}(A_\tau \cdot \dots \cdot A_1)^{1/\tau}, \quad (144)$$

as formally proved in Hössjer et al. (2014). It is straightforward to extend Theorem 1 by measuring time in units of  $\tau$ , so that the allele frequency Markov process has kernel  $P_1 \cdot \dots \cdot P_\tau$ . Then (144) equals the rate of fixation  $\lambda = \lambda_3(P_1 \cdot \dots \cdot P_\tau)^{1/\tau}$  of alleles in units of time step one.

Second, we have included two-sex models, defining subpopulations in terms of male and female gametes. It would also be of interest to define subpopulations in terms of individuals, as for an island model with diploid monoecious or dioecious individuals (Chesser et al. 1993; Wang 1997a, b). This would require some changes in the way the elements of  $A$  are characterized in terms of coalescence probabilities, requiring modifications of (34) and Theorem 3.

**Acknowledgments** The author was financially supported by the Gustafsson Foundation for Research in Natural Sciences and Medicine and the Swedish Research Council, contracts nr. 621-2008-4946 and 621-2013-4633. He thanks an associate editor and two anonymous reviewers for very helpful comments that considerably improved the manuscript, Nils Ryman for suggesting the source population example of Sect. 5 and for provision of valuable references, and Fredrik Olsson for the plots and numerical values of Sect. 6.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

## 8 Appendix

*Proof of Theorem 1* Formula (14) follows from the lower block diagonal decomposition (13) of  $\mathbf{P}$ , which implies that the spectrum of eigenvalues of  $\mathbf{P}$  is the union of the spectrum of eigenvalues of all  $\mathbf{P}_{ii}$ . Moreover, for each  $i \geq 3$  we have that  $\lambda_{\max}(\mathbf{P}_{ii}) < 1$ , since  $\mathbf{P}_{ii}$  has non-negative elements, row sums less or equal to one, with at least one row sum strictly less than one. In conjunction with (14), this implies  $\lambda_3 < 1$ .

In order to prove (15), we recall from (12) that  $\phi_2(\mathbf{x}) = \gamma \mathbf{x}$ . It follows from (10) that  $\{\phi_2(\mathbf{X}_t)\}$  is a martingale. With  $\tau = \min\{t; \mathbf{X}_t \in \mathcal{X}_1 \cup \mathcal{X}_2\}$  a stopping time, we can use the Optional Stopping Theorem (cf. Chapter 7 of [Grimmett and Stirzaker 2001](#)) to deduce

$$\begin{aligned} \lim_{t \rightarrow \infty} \mathbf{P}^t(\mathbf{x}, \mathbf{1}) &= P_{\mathbf{x}}(\mathbf{X}_{\tau} = \mathbf{1}) \\ &= E_{\mathbf{x}}(\phi_2(\mathbf{X}_{\tau})) \\ &= E_{\mathbf{x}}(\phi_2(\mathbf{X}_0)) \\ &= \phi_2(\mathbf{x}), \end{aligned} \quad (145)$$

and similarly  $\mathbf{P}^t(\mathbf{x}, \mathbf{0}) \rightarrow 1 - \phi_2(\mathbf{x}) = \phi_1(\mathbf{x})$ , for any  $\mathbf{x} \in \mathcal{X}$ . This establishes the leading two terms on the right hand side of (15). In order motivate the next term of order  $\lambda_3^t$ , we notice that  $\mathbf{P}_{kk}$  is non-negative, irreducible and aperiodic, and therefore has a unique largest eigenvalue  $\lambda_3$ . But since the maximum in (14) is attained uniquely for  $i = k$ ,  $\lambda_3$  must also be a simple eigenvalue for  $\mathbf{P}$ , so that  $|\lambda_i| < \lambda_3$  for  $i = 4, \dots, |\mathcal{X}|$ . In conjunction with (145), this implies

$$\mathbf{P}^t = \phi_1 \pi_1 + \phi_2 \pi_2 + \lambda^t \mathbf{R} + o(\lambda^t), \quad (146)$$

for some matrix  $\mathbf{R}$  as  $t \rightarrow \infty$ . Since  $\mathbf{P}$  has a lower triangular block decomposition by (11), so has  $\mathbf{P}^t = (\mathbf{P}_{ij}^{(t)})$  and  $\mathbf{R} = (\mathbf{R}_{ij})$ . Moreover,  $\mathbf{P}^t$  has non-negative elements and  $\phi_1 \pi_1 + \phi_2 \pi_2$  has all but its first two columns equal to zero, and therefore each  $\mathbf{R}_{ij}$  with  $j \geq 3$  is non-negative. Apply Perron-Frobenius Theorem to the non-negative, irreducible and aperiodic matrix  $\mathbf{P}_{kk}$  in order to deduce  $\mathbf{P}_{kk}^{(t)} = \lambda_3^t \phi_k \mathbf{q}_k + o(\lambda_3^t)$ , with  $\phi_k$  and  $\mathbf{q}_k$  the right and left eigenvectors for the leading eigenvalue  $\lambda_3$  with strictly positive components, normalized for instance so that  $\sum_{\mathbf{x} \in \mathcal{X}_k} \mathbf{q}_k(\mathbf{x}) = \sum_{\mathbf{x} \in \mathcal{X}_k} \phi_k(\mathbf{x}) \mathbf{q}_k(\mathbf{x}) = 1$ . This proves that  $\mathbf{R}_k = \phi_k \mathbf{q}_k$ , and since all other block diagonal  $\mathbf{P}_{ii}$  have leading eigenvalues smaller than  $\lambda_3$ , the corresponding matrices  $\mathbf{R}_{ii}$  must be zero.

The remaining submatrices  $\mathbf{R}_{ij}$  of  $\mathbf{R}$  ( $i > j \geq 1, i \geq 3$ ) can be computed as follows. We use (15) and the two recursions  $\mathbf{P}_{ij}^{(t+1)} = \sum_{l=j}^i \mathbf{P}_{il} \mathbf{P}_{lj}^{(t)}$  and  $\mathbf{P}_{ij}^{(t+1)} =$

$\sum_{l=j}^i \mathbf{P}_{il}^{(t)} \mathbf{P}_{lj}$  and let  $t \rightarrow \infty$ . After some computations this leads to

$$\begin{aligned} \mathbf{R}_{ij} &= (\lambda_3 \mathbf{I}_{|\mathcal{X}_i|} - \mathbf{P}_{ii})^{-1} \sum_{l=\max(j,3)}^{i-1} \mathbf{P}_{il} \mathbf{R}_{lj}, \quad i = j+1, \dots, n, \\ \mathbf{R}_{ij} &= \sum_{l=\max(j+1,3)}^i \mathbf{R}_{il} \mathbf{P}_{lj} (\lambda_3 \mathbf{I}_{|\mathcal{X}_j|} - \mathbf{P}_{jj})^{-1}, \quad j = i-1, \dots, 1, \end{aligned} \quad (147)$$

with  $\mathbf{I}_p$  an identity matrix of order  $p$ . The upper and lower recursions of (147) is applicable for  $i \neq k$  and  $j \neq k$ , and all  $\mathbf{R}_{ij}$  can be found from them.  $\square$

*Proof of Corollary 1* It follows from (15) and (17) that

$$\begin{aligned} E_\pi(\phi(X_t)) &= \pi \mathbf{P}^t \phi \\ &= (\pi \phi_1)(\pi_1 \phi) + (\pi \phi_2)(\pi_2 \phi) + \lambda_3^t \pi \mathbf{R} \phi + o(\lambda_3^t) \\ &= \lambda_3^t \sum_{\mathbf{x}, \mathbf{y} \in \mathcal{X}} \pi(\mathbf{x}) \mathbf{R}(\mathbf{x}, \mathbf{y}) \phi(\mathbf{y}) + o(\lambda_3^t) \\ &= \lambda_3^t \sum_{\mathbf{x} \in \mathcal{X}, \mathbf{y} \in \mathcal{X} \setminus (\mathcal{X}_1 \cup \mathcal{X}_2)} \pi(\mathbf{x}) \mathbf{R}(\mathbf{x}, \mathbf{y}) \phi(\mathbf{y}) + o(\lambda_3^t) \\ &\geq \lambda_3^t \sum_{\mathbf{x}, \mathbf{y} \in \mathcal{X}_k} \pi(\mathbf{x}) \mathbf{R}(\mathbf{x}, \mathbf{y}) \phi(\mathbf{y}) + o(\lambda_3^t) \\ &= \lambda_3^t \sum_{\mathbf{x}, \mathbf{y} \in \mathcal{X}_k} \pi(\mathbf{x}) \phi_k(\mathbf{x}) q_k(\mathbf{y}) \phi(\mathbf{y}) + o(\lambda_3^t) \\ &= \lambda_3^t \sum_{\mathbf{x} \in \mathcal{X}_k} \pi(\mathbf{x}) \phi_k(\mathbf{x}) \sum_{\mathbf{y} \in \mathcal{X}_k} q_k(\mathbf{y}) \phi(\mathbf{y}) + o(\lambda_3^t) \\ &> 0, \end{aligned}$$

where the last inequality holds if  $\pi(\mathcal{X}_k) > 0$ , since  $q_k$ ,  $\phi_k$  and  $\phi$  are strictly positive functions on  $\mathcal{X}_k$ .  $\square$

*Proof of Proposition 2* The proposition follows from

$$\begin{aligned} E(\phi_W(X_t)) &= 2 \sum_{i,j=1}^s W_{ij} E_\pi(X_{ti}(1 - X_{tj})) \\ &= \sum_{i,j=1}^s W_{ij} (E_\pi(X_{ti}(1 - X_{tj})) + E_\pi(X_{tj}(1 - X_{ti}))) \\ &= \sum_{i,j=1}^s W_{ij} H_{tij}, \end{aligned}$$

where in the second step we used the symmetry condition  $W_{ij} = W_{ji}$ , and in the last step the definition of  $H_{tij}$  in (23).  $\square$



*Proof of Theorem 2* Assume that all elements  $W_{ij}$  of  $\mathbf{W}$  are strictly positive, and that the symmetry condition  $W_{ji} = W_{ij}$  holds. Then  $\phi_{\mathbf{W}}$  satisfies (17), and it follows from (18) and Proposition 2 that

$$\mathbf{W}\mathbf{H}_t = C\lambda^t + o(\lambda^t) \text{ as } t \rightarrow \infty, \quad (148)$$

with  $C > 0$ . On the other hand we can use (25) repeatedly  $t$  times to deduce

$$\mathbf{W}\mathbf{H}_t = \mathbf{W}\mathbf{A}^t\mathbf{H}_0. \quad (149)$$

The block decomposition (28) implies that  $\lambda_{\max}(\mathbf{A}) = \max_{a=1,\dots,m} \lambda_{\max}(\mathbf{A}_{aa})$ , and since (148)–(149) hold for any vector  $\mathbf{W}$  with non-negative components and any admissible  $\mathbf{H}_0$ , (29) follows. In the sequel, we therefore write  $\lambda = \lambda_{\max}(\mathbf{A})$ .

If the maximum in (29) is attained for a unique  $1 \leq c \leq m$ , we lump  $\mathcal{I}_2 = \bar{\mathcal{C}}_1 \cup \bar{\mathcal{C}}_2 \cup \bar{\mathcal{C}}_3$  into (at most) three components  $\bar{\mathcal{C}}_1 = \cup_{a=1}^{c-1} \mathcal{C}_a$ ,  $\bar{\mathcal{C}}_2 = \mathcal{C}_c$  and  $\bar{\mathcal{C}}_3 = \cup_{a=c+1}^m \mathcal{C}_a$ . In particular, we put  $\bar{\mathcal{C}}_1 = \emptyset$  if  $c = 1$ , and  $\bar{\mathcal{C}}_3 = \emptyset$  if  $c = m$ . The corresponding block decomposition of  $\mathbf{A}$  is

$$\mathbf{A} = \begin{pmatrix} \bar{\mathbf{A}}_{11} & \mathbf{0} & \mathbf{0} \\ \bar{\mathbf{A}}_{21} & \bar{\mathbf{A}}_{22} & \mathbf{0} \\ \bar{\mathbf{A}}_{31} & \bar{\mathbf{A}}_{32} & \bar{\mathbf{A}}_{33} \end{pmatrix}. \quad (150)$$

It can be shown that matrices  $\mathbf{C}_{ab}$  exist for  $1 \leq b \leq a \leq 3$  such that

$$\mathbf{A}^t = \begin{pmatrix} \bar{\mathbf{A}}_{11}^{(t)} & \mathbf{0} & \mathbf{0} \\ \bar{\mathbf{A}}_{21}^{(t)} & \bar{\mathbf{A}}_{22}^{(t)} & \mathbf{0} \\ \bar{\mathbf{A}}_{31}^{(t)} & \bar{\mathbf{A}}_{32}^{(t)} & \bar{\mathbf{A}}_{33}^{(t)} \end{pmatrix} = \lambda^t \begin{pmatrix} \mathbf{C}_{11} & \mathbf{0} & \mathbf{0} \\ \mathbf{C}_{21} & \mathbf{C}_{22} & \mathbf{0} \\ \mathbf{C}_{31} & \mathbf{C}_{32} & \mathbf{C}_{33} \end{pmatrix} + o(\lambda^t) \quad (151)$$

as  $t \rightarrow \infty$ . We will identify the components of  $\mathbf{C} = (\mathbf{C}_{ab})$ , and in this process find the right and left eigenvectors  $\mathbf{r}$  and  $\boldsymbol{\rho}$  of  $\mathbf{A}$ .

Starting with the diagonal submatrices of  $\mathbf{C}$ , it follows that  $\mathbf{C}_{11} = \mathbf{C}_{33} = \mathbf{0}$ , since  $\lambda_{\max}(\bar{\mathbf{A}}_{aa}) < \lambda$  for  $a = 1, 3$ . On the other hand, since  $\bar{\mathbf{A}}_{22} = \mathbf{A}_{cc}$  is irreducible with  $\lambda_{\max}(\bar{\mathbf{A}}_{aa}) = \lambda$ , Perron–Frobenius Theorem and (149) imply that  $\lambda$  is a simple eigenvalue of  $\bar{\mathbf{A}}_{22}$  with periodicity 1. Thus we can find a right eigenvector  $\mathbf{r}_2 = (r_{2ij}; i, j \in \bar{\mathcal{C}}_2)'$ , and a left eigenvector  $\boldsymbol{\rho}_2 = (\rho_{2ij}; i, j \in \bar{\mathcal{C}}_2)$  of  $\bar{\mathbf{A}}_{22}$  with strictly positive components, normalized so that  $\sum_{ij \in \bar{\mathcal{C}}_2} \rho_{2ij} = \sum_{ij \in \bar{\mathcal{C}}_2} r_{2ij} \rho_{2ij} = 1$ , with

$$\mathbf{C}_{22} = \mathbf{r}_2 \boldsymbol{\rho}_2. \quad (152)$$

For the non-diagonal elements of  $\mathbf{C}$  we use the three recursions

$$\begin{aligned} \bar{\mathbf{A}}_{21}^{(t+1)} &= \bar{\mathbf{A}}_{21}^{(t)} \bar{\mathbf{A}}_{11} + \bar{\mathbf{A}}_{22}^{(t)} \bar{\mathbf{A}}_{21}, \\ \bar{\mathbf{A}}_{32}^{(t+1)} &= \bar{\mathbf{A}}_{32}^{(t)} \bar{\mathbf{A}}_{22} + \bar{\mathbf{A}}_{33}^{(t)} \bar{\mathbf{A}}_{32}, \\ \bar{\mathbf{A}}_{31}^{(t+1)} &= \bar{\mathbf{A}}_{32}^{(t)} \bar{\mathbf{A}}_{21} + \bar{\mathbf{A}}_{33}^{(t)} \bar{\mathbf{A}}_{31} + o(\lambda^t), \end{aligned} \quad (153)$$

insert (151) into (153), divide both sides of all three equations in (153) by  $\lambda^t$  and let  $t \rightarrow \infty$ . After some computations, this yields

$$\begin{aligned} C_{21} &= C_{22} \bar{A}_{21} (\lambda I_{|\bar{C}_1|} - \bar{A}_{11})^{-1}, \\ C_{32} &= (\lambda I_{|\bar{C}_3|} - \bar{A}_{33})^{-1} \bar{A}_{32} C_{22}, \\ C_{31} &= (\lambda I_{|\bar{C}_3|} - \bar{A}_{33})^{-1} \bar{A}_{32} C_{22} \bar{A}_{21} (\lambda I_{|\bar{C}_1|} - \bar{A}_{11})^{-1}, \end{aligned} \quad (154)$$

with  $I_p$  an identity matrix of order  $p$ . Finally, inserting (152) into (154), we find that  $C = r\rho = (\theta', r'_2, r'_3)'(\rho_1, \rho_2, \mathbf{0})$ , with

$$\begin{aligned} \rho_1 &= \rho_2 \bar{A}_{21} (\lambda I_{|\bar{C}_1|} - \bar{A}_{11})^{-1}, \\ r_3 &= (\lambda I_{|\bar{C}_3|} - \bar{A}_{33})^{-1} \bar{A}_{32} r_2. \end{aligned} \quad (155)$$

Both  $r_3$  and  $\rho_1$  must have non-negative elements, since  $r_2$  and  $\rho_2$  have strictly positive elements, all  $\bar{A}_{ab}$  are non-negative and  $\bar{A}_{11}$ ,  $\bar{A}_{33}$  have all their eigenvalues less than  $\lambda$ . Finally, since  $r_2$  and  $\rho_2$  are right and left eigenvectors of  $\bar{A}_{22}$  with eigenvalue  $\lambda$ , it follows after some computations from (155) and the block decomposition (150) of  $A$ , that  $r$  and  $\rho$  are right and left eigenvectors of  $A$  with eigenvalue  $\lambda$ .  $\square$

*Proof of Theorem 3* Recall that  $\mathcal{H}_{tij}$  in (22) is the probability that two genes at time  $t$  have different types of alleles when picked independently from subpopulations  $i$  and  $j$ , with replacement if  $i = j$ . Conditionally on  $\mathcal{B}_t$  and  $X_{t-1}$ , we compute the expected value of this probability by conditioning on the parental subpopulations  $k$  and  $l$  of  $i$  and  $j$ , and then take into account whether the two parental genes are identical or not. We find that

$$E(\mathcal{H}_{tij} | \mathcal{B}_t, X_{t-1}) = \left(1 - \frac{1}{2Nu_i}\right)^{\{i=j\}} \sum_{k,l} \mathcal{Q}_{tij,kl} \mathcal{H}_{t-1,kl} \left(\frac{1 - \mathcal{P}_{tijk}}{1 - \frac{1}{2Nu_k}}\right)^{\{k=l\}}, \quad (156)$$

since the probability of picking two different genes at time  $t$  is  $1 - 1/(2Nu_i)$  when  $i = j$  and 1 when  $i \neq j$ . Then the probability that the two parental genes from subpopulations  $k$  and  $l$  are different is 1 when  $k \neq l$  and  $1 - \mathcal{P}_{tijk}$  when  $k = l$ . In the former  $k \neq l$  case, the probability is  $\mathcal{H}_{t-1,kl}$  that the parental genes are different by state, and when  $k = l$ , the probability is  $\mathcal{H}_{t-1,kk}/(1 - (2Nu_k)^{-1})$  that the two parental genes are different by state.

We can express the coalescence probability  $p_{ijk}$  in (37) in terms of  $\mathcal{P}_{tijk}$  as

$$\begin{aligned}
 p_{ijk} &= P(T = 1 | I_0 = i, J_0 = j, I_1 = J_1 = k, T > 0) \\
 &= P(T = 1, I_1 = J_1 = k | I_0 = i, J_0 = j, T > 0) \\
 &\quad / P(I_1 = J_1 = k | I_0 = i, J_0 = j, T > 0) \\
 &= E(P(T = 1, I_1 = J_1 = k | I_0 = i, J_0 = j, T > 0, \mathcal{B}_t)) / Q_{ij, kk} \\
 &= E(P(T = 1 | I_1 = J_1 = k, I_0 = i, J_0 = j, T > 0, \mathcal{B}_t)) \\
 &\quad \cdot P(I_1 = J_1 = k | I_0 = i, J_0 = j, T > 0, \mathcal{B}_t)) / Q_{ij, kk} \\
 &= E(\mathcal{P}_{tijk} Q_{tij, kk}) / Q_{ij, kk}.
 \end{aligned} \tag{157}$$

Then average with respect to  $\mathcal{B}_t$  on the left and right hand sides of (156), use independence between  $\mathcal{B}_t$  and  $X_{t-1}$ , invoke (35), (41) and (157), to find that

$$\begin{aligned}
 E(\mathcal{H}_{tij} | X_{t-1}) &= \left(1 - \frac{1}{2Nu_i}\right)^{\{i=j\}} \sum_{k,l} Q_{ij,kl} \mathcal{H}_{t-1,kl} \left(\frac{1 - p_{ijk}}{1 - \frac{1}{2Nu_k}}\right)^{\{k=l\}} \\
 &= \sum_{k,l} A_{ij,kl} \mathcal{H}_{t-1,kl},
 \end{aligned} \tag{158}$$

in accordance with (40). As a next step we average both sides of (158) with respect to  $X_{t-1}$ , using starting distribution  $\pi$  for  $X_0$ , and get

$$\begin{aligned}
 H_{tij} &= E_\pi(E(\mathcal{H}_{tij} | X_{t-1})) \\
 &= \sum_{k,l} A_{ij,kl} E_\pi(\mathcal{H}_{t-1,kl}) \\
 &= \sum_{k,l} A_{ij,kl} H_{t-1,kl},
 \end{aligned} \tag{159}$$

which is equivalent to (25). To verify (39), we first compute

$$\begin{aligned}
 P\left(T = 1 | I_0 = i, J_0 = j, I_1 = J_1 = k, T > 0, \mathcal{B}_t, \{\mathbf{v}_{tkg}\}_{g=1}^{2Nu_k}\right) \\
 = \sum_{g=1}^{2Nu_k} v_{tkig} (v_{tkjg} - 1_{\{i=j\}}) / (2Nu_i \mathcal{B}_{tik} (2Nu_j \mathcal{B}_{tjk} - 1_{\{i=j\}}))
 \end{aligned} \tag{160}$$

and introduce the variables

$$\mathcal{V}_{tkij} = \begin{cases} E(v_{tki1}(v_{tki1} - 1) | \mathcal{B}_t), & \text{if } i = j, \\ E(v_{tki1} v_{tkj1} | \mathcal{B}_t), & \text{if } i \neq j, \end{cases} \tag{161}$$

which are conditional versions of  $V_{kij}$  in (38). Then average with respect to the exchangeable random vectors  $\{\mathbf{v}_{tkg}\}_{g=1}^{2Nu_k}$  in (160) to deduce that

$$\begin{aligned}\mathcal{P}_{tijk} &= P(T = 1 | I_0 = i, J_0 = j, I_1 = J_1 = k, T > 0, \mathcal{B}_t) \\ &= 2Nu_k \mathcal{V}_{tkij} / (2Nu_i \mathcal{B}_{tik} \cdot (2Nu_j \mathcal{B}_{tjk} - 1_{\{i=j\}})) \\ &= u_k \mathcal{V}_{tkij} / \left( 2Nu_i \mathcal{B}_{tik} u_j \mathcal{B}_{tjk} \left( 1 - 1_{\{i=j\}} (2Nu_i \mathcal{B}_{tik})^{-1} \right) \right). \quad (162)\end{aligned}$$

We can rewrite (34) as

$$\mathcal{Q}_{tij,kl} = \mathcal{B}_{tik} \mathcal{B}_{tjl} \left( 1 - \frac{1}{2Nu_i \mathcal{B}_{tik}} \right)^{\{i=j, k=l\}} \left( \frac{1}{1 - \frac{1}{2Nu_i}} \right)^{\{i=j\}},$$

and taking the product of the last two displayed equations, we find that

$$\mathcal{P}_{tijk} \mathcal{Q}_{tij,kk} = \left( \frac{1}{1 - \frac{1}{2Nu_i}} \right)^{\{i=j\}} \frac{\mathcal{V}_{tkij} u_k}{2Nu_i u_j}, \quad (163)$$

Formula (39) then follows from (157) by averaging with respect to  $\mathcal{B}_t$  in (163), using that  $E(\mathcal{V}_{tkij}) = V_{kij}$ , and finally dividing by  $\mathcal{Q}_{ij,kk}$ .  $\square$

*Proof of Corollary 3* Define  $\phi_3$  as in (42), and the gene diversity vector  $\mathcal{H}(\mathbf{x}) = \text{vec} \left( (x_i(1-x_j) + x_j(1-x_i))_{i,j=1}^s \right)'$  of length  $s^2$  obtained from an allele frequency vector  $\mathbf{x}$  of length  $s$ . Then

$$\begin{aligned}(P\phi_3)(\mathbf{x}) &= \sum_{\mathbf{y}} P(\mathbf{x}, \mathbf{y}) \phi_3(\mathbf{y}) \\ &= \sum_{\mathbf{y}} P(\mathbf{x}, \mathbf{y}) \sum_{ij} \rho_{ij} (y_i(1-y_j) + y_j(1-y_i)) \\ &= \sum_{ij} \rho_{ij} \sum_{\mathbf{y}} P(\mathbf{x}, \mathbf{y}) (y_i(1-y_j) + y_j(1-y_i)) \\ &= \sum_{ij} \rho_{ij} E(\mathcal{H}_{tij} | X_{t-1} = \mathbf{x}) \\ &= \sum_{ij} \rho_{ij} (A\mathcal{H}(\mathbf{x}))_{ij} \\ &= \boldsymbol{\rho} A \mathcal{H}(\mathbf{x}) \\ &= \lambda \boldsymbol{\rho} \mathcal{H}(\mathbf{x}) \\ &= \lambda \sum_{ij} \rho_{ij} (x_i(1-x_j) + x_j(1-x_i)) \\ &= \lambda \phi_3(\mathbf{x}).\end{aligned}$$

In the fifth step we used (40), and in the seventh step that  $\rho$  is a left eigenvector of  $A$  with eigenvalue  $\lambda$ . Formula (43) is proved in the same way as the corresponding result for the right eigenvector  $r$  of  $A$  in Theorem 2.  $\square$

**Verifying (63), (64) and (65)** We first insert (7) and the expression for  $Q_{ij,kl}$  in (51) into (39), and find that  $p_{ijk} = [V_{kij}/(M_{ki}M_{kj})]/(2Nu_k - 1_{\{i=j\}}u_k/u_i)$ . Since  $E(v_{tki1}) = M_{ki}$ , it follows from (38) that

$$p_{ijk} = \begin{cases} [1 + (\text{Var}(v_{tki1}) - M_{ki})/M_{ki}^2]/(2Nu_k - u_k/u_i), & i = j, \\ [1 + \text{Cov}(v_{tki1}, v_{tkj1})/(M_{ki}M_{kj})]/(2Nu_k), & i \neq j. \end{cases} \quad (164)$$

In order to compute  $\text{Var}(v_{tki1})$  and  $\text{Cov}(v_{tki1}, v_{tkj1})$  for  $k = 1$ , the subpopulation of grandpaternally inherited alleles of the fathers of time  $t - 1$ , let  $\zeta_{tmmg} = v_{t11g} + v_{t21g'}$  and  $\zeta_{tmfg} = v_{t13g} + v_{t23g'}$  be the total number of children that are males and females respectively, of the male in time  $t - 1$  whose paternally and maternally inherited genes have been assigned number  $g$  and  $g' = g'(g)$  from  $1, \dots, N_m = 2Nu_1$ . (Due to the convention (9) that the first  $2Nu_k X_{t-1,k}$  genes of subpopulation  $k$  have the specified allele, we cannot assume  $g' = g$ .) Due to exchangeability,  $\tau_{mm}^2 = \text{Var}(\zeta_{tmm1})$ ,  $\tau_{mf}^2 = \text{Var}(\zeta_{tmf1})$  and  $\tau_{mm,mf} = \text{Cov}(\zeta_{tmm1}, \zeta_{tmf1})$ .

It follows from Mendel's law of inheritance that either the paternally or maternally inherited gamete is passed on to the offspring, with equal probability 0.5, independently between matings. Hence  $v_{t11g}|\zeta_{tmmg} \sim \text{Bin}(\zeta_{tmmg}, 0.5)$  and  $v_{t13g}|\zeta_{tmfg} \sim \text{Bin}(\zeta_{tmfg}, 0.5)$ , which implies

$$\begin{aligned} \text{Var}(v_{t11g}) &= \text{Var}(E(v_{t11g}|\zeta_{tmmg})) + E(\text{Var}(v_{t11g}|\zeta_{tmmg})) \\ &= \text{Var}\left(\frac{1}{2}\zeta_{tmmg}\right) + E\left(\frac{1}{4}\zeta_{tmmg}\right) \\ &= \frac{1}{4}\text{Var}(\zeta_{tmmg}) + \frac{1}{2}E(v_{t11g}) \\ &= \frac{1}{4}\tau_{mm}^2 + \frac{1}{2}M_{11}, \end{aligned}$$

and analogously  $\text{Var}(v_{t13g}) = \tau_{mf}^2/4 + M_{13}/2$  and  $\text{Cov}(v_{t11g}, v_{t13g}) = \tau_{mm,mf}/4$ . Inserting the last three formulas and (62) into (164), we see that  $p_{111}$ ,  $p_{331}$  and  $p_{131}$  simplify to (63), (64) and (65) respectively.  $\square$

*Proof of Theorem 5* Formulas (78) and (79) follow easily from a first order Taylor expansion of (73) with respect to  $\varepsilon$  at 0, making use of (75) and (77). Next we apply Theorem 4. We want to show that (80) follows from (69), with  $\varepsilon$  as in (74) and that

(81) is obtained by inserting (79) into (72). Indeed,

$$\begin{aligned} C &= \sum_{ijk} \rho_{ij}(0) \frac{\sigma_{ijk} Q_{ij,kk}(0)}{u_k^\beta} - \sum_{ijkl} \rho_{ij}(0) \dot{Q}_{ij,kl} \\ &\quad + 1_{\{\beta=1\}} \left( \sum_{ikl} \rho_{ii}(0) \frac{Q_{ii,kl}(0)}{u_i} - \sum_{ijk} \rho_{ij}(0) \frac{Q_{ij,kk}(0)}{u_k} \right) \\ &= \sum_{ijk} \rho_{ij}(0) \frac{\sigma_{ijk} Q_{ij,kk}(0)}{u_k^\beta}, \end{aligned}$$

where in the last step we employed that  $\rho(0)$  is a left eigenvector of  $Q(0)$  with eigenvalue 1, and moreover that  $\sum_{kl} Q_{ij,kl}(\varepsilon) = 1$  for all  $\varepsilon \geq 0$ , which implies  $\sum_{kl} Q_{ij,kl}(0) = 1$  and  $\sum_{kl} \dot{Q}_{ij,kl} = 0$ .  $\square$

*Proof of Corollary 5 and 6* Corollary 4 implies that  $Q_{ij,kl}(0) = B_{ik} B_{jl}$  in Corollary 6, so that

$$\sum_{ij} \gamma_i \gamma_j Q_{ij,kl}(0) = \sum_{ij} \gamma_i \gamma_j B_{ik} B_{jl} = \sum_i \gamma_i B_{ik} \sum_j \gamma_j B_{il} = \gamma_k \gamma_l$$

follows from (4), and  $\gamma \otimes \gamma = (\gamma_i \gamma_j)$  is a left eigenvector of  $Q(0)$  with eigenvalue 1. This proves (89), and insertion into (81) yields (90). We then use (7) and (76) in order to prove (91), since

$$\sigma_{ijk} = \lim_{N \rightarrow \infty} \frac{V_{kij} u_k^2}{(Nu_k)^{1-\beta} u_i u_j (M_{ki} u_k / u_i) (M_{kj} u_k / u_j)}.$$

In Corollary 5, it follows from Corollary 4 that  $Q_{ii,kl}(0) = 1_{\{k=l\}} B_{ik}$ , and

$$\sum_{ij} \gamma_i 1_{\{i=j\}} Q_{ij,kl}(0) = \sum_i \gamma_i 1_{\{k=l\}} B_{ik} = 1_{\{k=l\}} \gamma_k,$$

because of (4). This proves (86), which inserted into (81) yields (87). Since

$$\begin{aligned} V_{kii} &= E(v_{tki} 1(v_{tki} - 1)) \\ &= P(K_{ti} = k) E(v_{tki} 1(v_{tki} - 1) | K_{ti} = k) \\ &= B_{ik} \bar{V}_{kii} \end{aligned}$$

and  $Q_{ii,kk}(0) = B_{ik}$ , we get  $\sigma_{iik} = \lim_{N \rightarrow \infty} (B_{ik} \bar{V}_{kii} u_k^2) / ((Nu_k)^{1-\beta} u_i^2 B_{ik})$  from (76), thereby proving (88).  $\square$

**Motivation of Lemma 1.** We first sketch a proof of (110), using similar calculations as in Möhle (1998a). By the definition of an equilibrium distribution of a Markov

chain,  $\mathbf{B}(\varepsilon)^t \rightarrow \mathbf{1}\boldsymbol{\gamma}(\varepsilon)$  as  $t \rightarrow \infty$ , where  $\mathbf{1}$  is a column vector of ones of length  $s$ . Hence we will study  $\mathbf{B}(\varepsilon)^t$  for large  $t$  and use the approximation

$$\mathbf{B}(\varepsilon)^t = (\mathbf{B}(0) + \varepsilon \dot{\mathbf{B}})^t \approx \sum_{r=0}^t \binom{t}{r} \varepsilon^r (\mathbf{B}(0)^\infty \dot{\mathbf{B}} \mathbf{B}(0)^\infty)^r, \quad (165)$$

which can be shown to be accurate in the limit  $\varepsilon \rightarrow 0$ , with

$$\mathbf{B}(0)^\infty = \begin{pmatrix} \mathbf{B}_{11}(0)^\infty & \dots & 0 \\ & \ddots & \\ 0 & \dots & \mathbf{B}_{mm}(0)^\infty \end{pmatrix} = \begin{pmatrix} \mathbf{1}_1 \boldsymbol{\gamma}_1 & \dots & 0 \\ & \ddots & \\ 0 & \dots & \mathbf{1}_m \boldsymbol{\gamma}_m \end{pmatrix}$$

a block diagonal matrix and  $\mathbf{1}_d = (1, \dots, 1)'$  a column vector of  $|\mathcal{I}(d)|$  ones. It follows from (108) and  $\boldsymbol{\gamma}_d = \boldsymbol{\gamma}_d \mathbf{B}_{dd}(0)$ , that

$$(\mathbf{B}(0)^\infty \dot{\mathbf{B}} \mathbf{B}(0)^\infty)^r = \begin{pmatrix} G_{11}^{(r)} \mathbf{1}_1 \boldsymbol{\gamma}_1 & \dots & G_{1m}^{(r)} \mathbf{1}_1 \boldsymbol{\gamma}_m \\ \vdots & & \vdots \\ G_{m1}^{(r)} \mathbf{1}_m \boldsymbol{\gamma}_1 & \dots & G_{mm}^{(r)} \mathbf{1}_m \boldsymbol{\gamma}_m \end{pmatrix},$$

where  $\mathbf{G}^r = (G_{ab}^{(r)})$ . Putting  $t = x/\varepsilon$  in (165), we notice that

$$\mathbf{B}(\varepsilon)^{x/\varepsilon} \approx \begin{pmatrix} \exp(x\mathbf{G})_{11} \mathbf{1}_1 \boldsymbol{\gamma}_1 & \dots & \exp(x\mathbf{G})_{1m} \mathbf{1}_1 \boldsymbol{\gamma}_m \\ \vdots & & \vdots \\ \exp(x\mathbf{G})_{m1} \mathbf{1}_m \boldsymbol{\gamma}_1 & \dots & \exp(x\mathbf{G})_{mm} \mathbf{1}_m \boldsymbol{\gamma}_m \end{pmatrix}$$

where

$$\exp(x\mathbf{G}) = \sum_{r=0}^{\infty} \frac{x^r}{r!} \mathbf{G}^r \xrightarrow{x \rightarrow \infty} \begin{pmatrix} \theta_1 & \dots & \theta_m \\ \vdots & & \vdots \\ \theta_1 & \dots & \theta_m \end{pmatrix},$$

since  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)$  is an equilibrium distribution of the continuous time Markov process with infinitesimal generator  $\mathbf{G}$ . Combining the last two displayed equations, we find for large  $x$  and small  $\varepsilon$  that

$$\mathbf{B}(\varepsilon)^{x/\varepsilon} \approx \begin{pmatrix} \theta_1 \mathbf{1}_1 \boldsymbol{\gamma}_1 & \dots & \theta_m \mathbf{1}_1 \boldsymbol{\gamma}_m \\ \vdots & & \vdots \\ \theta_1 \mathbf{1}_m \boldsymbol{\gamma}_1 & \dots & \theta_m \mathbf{1}_m \boldsymbol{\gamma}_m \end{pmatrix} = \mathbf{1}\boldsymbol{\gamma}(0),$$

where  $\boldsymbol{\gamma}(0) = (\gamma_i(0))$ . We use (106), (108) and (110) to prove (111);

$$\begin{aligned} B(\varepsilon) &= \sum_{d=1}^m \sum_{i \in \mathcal{I}(d)} \gamma_i(\varepsilon) \sum_{k \notin \mathcal{I}(d)} B_{ik}(\varepsilon) \\ &= \varepsilon \sum_{d=1}^m \sum_{i \in \mathcal{I}(d)} \gamma_i(\varepsilon) \sum_{k \notin \mathcal{I}(d)} \dot{B}_{ik} \\ &= \varepsilon \sum_{d=1}^m \sum_{i \in \mathcal{I}(d)} \gamma_i(0) \sum_{k \notin \mathcal{I}(d)} \dot{B}_{ik} + o(\varepsilon) \\ &= -\varepsilon \sum_{d=1}^m \sum_{i, k \in \mathcal{I}(d)} \gamma_i(0) \dot{B}_{ik} + o(\varepsilon), \end{aligned}$$

since the row sums of  $\dot{B}$  are zero. Finally, (112) follows from (105), since

$$\begin{aligned} M(\varepsilon) &= \sum_{d=1}^m \sum_{k \in \mathcal{I}(d)} u_k(\varepsilon) \sum_{i \notin \mathcal{I}(d)} M_{ki}(\varepsilon) \\ &= \varepsilon \sum_{d=1}^m \sum_{k \in \mathcal{I}(d)} u_k(0) \sum_{i \notin \mathcal{I}(d)} \dot{M}_{ki} + o(\varepsilon). \end{aligned}$$

□

**Proving (133).** By (105) and (132), the off-diagonal block elements of  $\dot{B}$  are

$$\begin{aligned} \dot{B}_{1,z+k} &= c_1 B_{z+1,z+k}(0) u_{z+1}, & k &= 1, \dots, z, \\ \dot{B}_{k+1,z+k} &= c_{k+1} u_{z+k+1}, & k &= 1, \dots, z-1, \\ \dot{B}_{z+1,k} &= c_1 B_{1k}(0) u_1, & k &= 1, \dots, z, \\ \dot{B}_{z+k+1,k} &= c_{k+1} u_{k+1}, & k &= 1, \dots, z-1. \end{aligned}$$

where  $B_{1k}(0)$  is the probability that the parent of a newborn in deme 1 originates from age class  $k$  of that deme when there is no migration, and similarly  $B_{z+1,z+k}(0)$  is the probability that the parent of a newborn in deme 2 has a parent from age class  $k$  of that deme. Invoking (108), the elements of  $G$  are

$$\begin{aligned} G_{12} = -G_{11} &= \sum_{i=1}^z \gamma_{1i} \sum_{k=1}^z \dot{B}_{i,z+k} = \sum_{i=1}^z \gamma_{1i} c_i u_{z+i}, \\ G_{21} = -G_{22} &= \sum_{i=1}^z \gamma_{2i} \sum_{k=1}^z \dot{B}_{z+i,k} = \sum_{i=1}^z \gamma_{2i} c_i u_i. \end{aligned}$$



Use (119) to deduce that  $\dot{\mathbf{A}} = (\dot{\Lambda}_{\alpha,\beta})_{\alpha,\beta=12,21}$  is a diagonal matrix of order  $v = 2(2-1) = 2$ , with both diagonal elements equal to  $G_{11} + G_{22}$ . Therefore

$$\lambda_{\max}(\dot{\mathbf{A}}) = G_{11} + G_{22} = - \sum_{i=1}^z c_i (\gamma_{1i} u_{z+i} + \gamma_{2i} u_i). \quad (166)$$

Because of (8) and (132), the numerator of (122) equals

$$\begin{aligned} & \sum_{k=1}^z u_k \sum_{i=1}^z \dot{M}_{k,z+i} + \sum_{k=1}^z u_{z+k} \sum_{i=1}^z \dot{M}_{z+k,i} \\ &= \sum_{k=1}^z u_k (c_1 M_{k1} u_{z+1} + c_{k+1} M_{k,k+1} u_{z+k+1} 1_{\{k < z\}}) \\ & \quad + \sum_{k=1}^z u_{z+k} (c_1 M_{z+k,z+1} u_1 + c_{k+1} M_{z+k,z+k+1} u_{k+1} 1_{\{k < z\}}) \\ &= 2 \sum_{k=1}^z c_k u_k u_{z+k}. \end{aligned} \quad (167)$$

Inserting (166) and (167) into (122), we arrive at (133).  $\square$

## References

- Brockwell PJ, Davis RA (1991) Time series: theory and methods, 2nd edn. Springer, New York
- Caballero A (1994) Developments in the prediction of effective population size. *Heredity* 73:657–679
- Caballero A (1995) On the effective size of populations with separate sexes, with particular reference to sex-linked genes. *Genetics* 139:1007–1011
- Caballero A, Hill WG (1992) A note on the inbreeding effective population size. *Evolution* 46(6):1969–1972
- Cannings C (1974) The latent roots of certain Markov chains arising in genetics: a new approach I. Haploid models. *Adv Appl Prob* 6:260–290
- Caswell H (2001) Matrix population models, 2nd edn. Sinauer, Sunderland
- Charlesworth B (2009) Effective population size and patterns of molecular evolution and variation. *Nat Rev Genet* 10:195–205
- Chesser RK, Rhodes OE, Sugg DW, Schnabel A (1993) Effective sizes for subdivided populations. *Genetics* 135:1221–1232
- Collet P, Martinez S (2013) Quasi stationary distributions, Markov chains, diffusions and dynamical systems. Springer, Berlin
- Cox DR, Miller HD (1965) The theory of stochastic processes. Methuen & Co Ltd, London
- Crow JF (1954) Breeding structure of populations II. Effective population number. In: *Statistics and Mathematics in Biology*. Iowas State Coll. Press, Ames, Iowa, pp 543–556
- Crow JF, Denniston C (1988) Inbreeding and variance effective population sizes. *Evolution* 42(3):482–495
- Darroch JN, Seneta E (1965) On quasi-stationary distributions in absorbing discrete-time finite Markov chains. *J Appl Prob* 2:88–100
- Durrett R (2008) Probability models for DNA sequence evolution, 2nd edn. Springer, New York
- Eldon B, Wakeley J (2006) Coalescent process when the distribution of offspring number among individuals is highly skewed. *Genetics* 172:2621–2633
- Eldon B, Wakeley J (2009) Coalescence times and  $F_{ST}$  under a skewed offspring distribution among individuals in a population. *Genetics* 181:615–629
- Ewens WJ (1982) On the concept of effective population size. *Theor Popul Biol* 21:373–378

- Ewens WJ (1989) The effective population size in the presence of catastrophes. In: Feldman M (ed) *Mathematical evolutionary theory*. Princeton University, Princeton, p 9
- Ewens WJ (2004) *Mathematical population genetics. I. Theoretical introduction*, 2nd edn. Springer, New York
- Feller W (1951) Diffusion processes in genetics. In: Neyman J (ed) *Proceedings of 2nd Berkeley symposium on Math. Stat. Prob.* University of California Press, Berkeley, pp 227–246
- Felsenstein J (1971) Inbreeding and variance effective numbers in populations with overlapping generations. *Genetics* 68:581–597
- Felsenstein J (1982) How can we infer geography and history from gene frequencies? *J Theor Biol* 96:9–20
- Fisher RA (1958) *The genetical theory of natural selection*, 2nd edn. Dover, New York
- Friswell MI (1996) The derivatives of repeated eigenvalues and their associated eigenvectors. *Trans ASME* 118:390–397
- Grimmett G, Stirzaker D (2001) *Probability and random processes*, 3rd edn. Oxford University Press, Oxford
- Herbots HM (1997) The structured coalescent. In: Donnelly P, Tavaré S (eds) *Progress in population genetics and human evolution*. Springer, New York, pp 231–255
- Hill WG (1972) Effective size of populations with overlapping generations. *Theor Popul Biol* 3:278–289
- Hill WG (1979) A note of effective population size with overlapping generations. *Genetics* 92:317–322
- Horn RA, Johnson CR (1985) *Matrix analysis*. Cambridge University Press, Cambridge
- Hössjer O (2011) Coalescence theory for a general class of structured populations with fast migration. *Adv Appl Prob* 43(4):1027–1047
- Hössjer O, Olsson F, Laikre L, Ryman N (2014) A general framework for modeling short and long term patterns of genetic differentiation and effective size of subdivided populations. Under revision for *Mathematical Biosciences*
- Hössjer O, Ryman N (2014) Quasi equilibrium, variance effective population size and fixation index for models with spatial structure. *J Math Biol*
- Jagers P, Sagitov S (2004) Convergence to the coalescent in populations of substantially varying size. *J Appl Prob* 41:368–378
- Kaj I, Krone SM, Lascoux M (2001) Coalescent theory for seed bank models. *J Appl Prob* 38:285–301
- Kaj I, Krone SM (2003) The coalescent process in a population of varying size. *J Appl Prob* 40:33–48
- Karlin S (1968) Rares of approach to homozygosity for finite stochastic models with variable population size. *Am Nat* 102:443–455
- Karlin S, McGregor J (1965) Direct product branching processes and related induced Markov chains I. Calculations of rates of approach to homozygosity. Bernoulli, Beys, Laplace Anniversary volume. Springer, Berlin-Heidelberg-New York, pp 11–145
- Kimura M (1953) 'Stepping stone' model of population. *Ann Rep Natl Inst Genet Jpn* 3:62–63
- Kimura M (1957) Some problems of stochastic processes in genetics. *Ann Math Statist* 28:882–901
- Kimura M, Weiss GH (1964) The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics* 61:763–771
- Kingman JFC (1982) The coalescent. *Stoch Proc Appl* 13:235–248
- Li W-H (1976) Effect of migration on genetic distance. *Am Nat* 110:841–847
- Malécot G (1951) Un traitement stochastique des problèmes linéaires (mutation, linkage, migration) en génétique de populations. *Annales de l'Université de Lyon A* 14:79–117
- Maruyama T (1970a) On the rate of decrease of heterozygosity in circular stepping stone models of populations. *Theor Popul Biol* 1:101–119
- Maruyama T (1970b) Effective number of alleles in subdivided populations. *Theor Popul Biol* 1:273–306
- Möhle M (1998a) A convergence theorem for Markov chains arising in population genetics and the coalescent with selfing. *Adv Appl Prob* 30:493–512
- Möhle M (1998b) Coalescent results for two-sex population models. *Adv Appl Prob* 30:513–520
- Möhle M (2000) Ancestral processes in population genetics—the coalescent. *J Theor Biol* 204:629–638
- Nagylaki T (1980) The strong-migration limit in geographically structured populations. *J Math Biol* 9:101–114
- Nagylaki T (1995) The inbreeding effective population number in dioecious populations. *Genetics* 139:473–485
- Nagylaki T (2000) Geographical invariance and the strong-migration limit in subdivided populations. *J Math Biol* 41:123–142
- Nei M (1973) Analysis of gene diversity in subdivided populations. *Proc Natl Acad Sci USA* 70:3321–3323

- Nei M (1987) Molecular evolutionary genetics. Columbia University Press, Columbia
- Nordborg M, Krone S (2002) Separation of time scales and convergence to the coalescent in structured populations. In: Slatkin M, Veuille M (eds) Modern development in theoretical population genetics. Oxford University Press, Oxford, pp 194–232
- Norris JR (2008) Markov chains. Cambridge University Press, Cambridge
- Notohara M (1990) The coalescent and the genealogical process in geographically structured populations. *J Math Biol* 29:59–75
- Notohara M (1993) The strong-migration limit for the genealogical process of geographically structured populations. *J Math Biol* 31:115–122
- Orive ME (1993) Effective population size in organisms with complex life-histories. *Theor Popul Biol* 44:316–340
- Pitman J (1999) Coalescents with multiple collisions. *Ann Probab* 27:1870–1902
- Pollak E (1980) Effective population numbers and mean time to extinction in dioecious populations with overlapping generations. *Math Biosci* 52:1–25
- Pollak E (2002) Eigenvalue effective population numbers for populations that vary cyclically in size. *Math Biosci* 177(178):11–24
- Pollak E (2010) Coalescent theory for a monoecious random mating population with a varying size. *J Appl Prob* 47:41–57
- Pollak E (2011) Coalescent theory for age-structured random mating populations with two sexes. *Math Biosci* 233:126–134
- Ryman N, Leimar O (2008) Effect of mutation on genetic differentiation among nonequilibrium populations. *Evolution* 62(9):2250–2259
- Sagitov S (1999) The general coalescent with asynchronous mergers of ancestral lines. *J Appl Probab* 36:1116–1125
- Sagitov S, Jagers P (2005) The coalescent effective size of age-structured populations. *Ann Appl Probab* 15(3):1778–1797
- Sampson KY (2006) Structured coalescent with nonconservative migration. *J Appl Prob* 43:351–362
- Sano A, Shimizu A, Iizuka M (2004) Coalescent process with fluctuating population size and its effective size. *Theor Popul Biol* 65:39–48
- Sawyer S (1976) Results for the stepping stone model for migration in population genetics. *Ann Prob* 4:699–728
- Sjödén P, Kaj I, Krone S, Lascoux M, Nordborg M (2005) On the meaning and existence of an effective population size. *Genetics* 169:1061–1070
- Slatkin M (1991) Inbreeding coefficients and coalescent times. *Genet Res Camb* 58:167–175
- Sved JA, Latter BDH (1977) Migration and mutation in stochastic models of gene frequency change. *J Math Biol* 5:61–73
- Tufto J, Engen S, Hindar K (1996) Inferring patterns of migration from gene frequencies under equilibrium conditions. *Genetics* 144:1911–1921
- Tufto J, Hindar K (2003) Effective size in management and conservation of subdivided populations. *J Theor Popul Biol* 222:273–281
- van der Aa NP, Ter Morsche HG, Mattheij RRM (2007) Computation of eigenvalue and eigenvector derivatives for a general complex-valued eigensystem. *Electron J Linear Algebra* 16:300–314
- Wakeley J (1998) Segregating sites in Wright's island model. *Theor Popul Biol* 53:166–174
- Wakeley J, Sargsyan O (2009) Extensions of the coalescent effective population size. *Genetics* 181:341–345
- Wang J (1997a) Effective size and  $F$ -statistics of subdivided populations. I. Monoecious species with partial selfing. *Genetics* 146:1453–1463
- Wang J (1997b) Effective size and  $F$ -statistics of subdivided populations. II. Dioecious species. *Genetics* 146:1465–1474
- Wang J, Caballero A (1999) Developments in predicting the effective size of subdivided populations. *Heredity* 82:212–226
- Wang Y, Pollak E (2000a) The effective number of a population that varies cyclically in size I. Discrete generations. *Math Biosci* 166:1–21
- Wang Y, Pollak E (2000b) The effective number of a population that varies cyclically in size. II. Overlapping generations. *Math Biosci* 179:161–181
- Waples RS (2002) Definition and estimation of effective population size in the conservation of endangered species. In: Beissinger SR, McCullogh DR (eds) Populations viability analysis. The University of Chicago Press, Chicago

- Whitlock MC, Barton NH (1997) The effective size of a subdivided population. *Genetics* 145:427–441
- Wright S (1938) Size of population and breeding structure in relation to evolution. *Science* 87:430–431
- Wright S (1931) Evolution in Mendelian populations. *Genetics* 16:97–159
- Wright S (1943) Isolation by distance. *Genetics* 28:114–138